

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2002, 09:22:17 ; Search time 10.85 Seconds

(Without alignments)
780.606 Million cell updates/sec

Title: US-09-445-480a-2

Perfect score: 1287
Sequence: 1 MALSQNAKESKGFVYMIWV.....VDEGVKVPSPVRFKRPYC 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1089	84.6	199	1	EOST_ACTEO
2	295.5	23.0	2766	1	THYG_MOUSE
3	295	22.9	2769	1	THYG_BOVIN
4	294.5	22.9	2768	1	THYG_RAT
5	272	21.1	2768	1	THYG_HUMAN
6	245.5	19.1	1375	1	NID2_HUMAN
7	244	19.0	844	1	SAX_RANCA
8	236	18.3	1403	1	NID2_MOUSE
9	156	12.1	279	1	HG2A_MOUSE
10	154	12.0	1245	1	NID2_MOUSE
11	153	11.9	280	1	HG2A_RAT
12	153	11.9	1247	1	NID2_HUMAN
13	150	11.7	296	1	HG2A_HUMAN
14	132.5	10.3	258	1	IBP4_BOVIN
15	129.5	10.1	254	1	IBP4_MOUSE
16	128.5	10.0	254	1	IBP4_RAT
17	124.5	9.7	237	1	IBP4_SHEEP
18	122.5	9.5	258	1	IBP4_HUMAN
19	122.5	9.5	271	1	IBP5_PIG
20	122.5	9.5	272	1	IBP5_HUMAN
21	120.5	9.4	271	1	IBP5_MOUSE
22	118.5	9.2	271	1	IBP5_RAT
23	114	8.9	263	1	IBP1_BOVIN
24	114	8.9	1786	1	IBP1_MOUSE
25	113.5	8.8	240	1	IBP6_HUMAN
26	112	8.7	1786	1	IBP1_HUMAN
27	110	8.5	266	1	IBP3_PIG
28	110	8.5	291	1	IBP3_BOVIN
29	109.5	8.5	259	1	IBP1_HUMAN
30	109.5	8.5	4544	1	IBP1_HUMAN
31	109	8.5	291	1	IBP3_MOUSE
32	108.5	8.4	292	1	IBP3_RAT
33	107	8.3	226	1	IBP6_MOUSE

34	107	8.3	238	1	IBP6_MOUSE	P47880 mus musculus
35	107	8.3	291	1	IBP3_HUMAN	P17936 homo sapien
36	107	8.3	305	1	IBP2_MOUSE	P47877 mus musculus
37	107	8.3	1429	1	L112_CAEEL	P14585 caenorhabdi
38	106	8.2	272	1	IBP1_RAT	P21743 rattus norv
39	105.5	8.2	863	1	LDVR_CHICK	P98165 gallus gall
40	104	8.1	810	1	IBP1_HUMAN	O92832 homo sapien
41	103.5	8.0	304	1	IBP2_HUMAN	P12843 rattus norv
42	103	8.0	272	1	IBP1_MOUSE	P47876 mus musculus
43	102	7.9	311	1	IBP2_CHICK	P49705 gallus gall
44	100.5	7.8	314	1	G732_HUMAN	P16422 homo sapien
45	100	7.8	1746	1	TENN_PIG	O29116 sus scrofa

ALIGNMENTS

RESULT	ID	EOST_ACTEO	STANDARD	PRT	199 AA.
AC	P81439				
DT	15-DEC-1998 (Rel. 37, Created)				
DT	15-DEC-1998 (Rel. 37, Last sequence update)				
DT	15-DEC-1998 (Rel. 37, Last annotation update)				
DE	EQUISTATIN.				
OS	Actinia equina (European sea anemone).				
OC	Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;				
OC	Nymphaeae; Actinidae; Actinia.				
OX	NCBI_TaxID=6106;				
RN	[1]				
RP	SEQUENCE.				
RX	MEDLINE=97298102; PubMed=9153250;				
RA	Lenarcic B., Ritonja A., Strukelj B., Turk B., Turk V.,				
RT	"Equislatin, a new inhibitor of cysteine proteinases from Actinia				
RT	equina, is structurally related to thyroglobulin type-1 domain.";				
RL	J. Biol. Chem. 272:13899-13903(1997).				
RN	[2]				
RP	REVIEWS.				
RA	Lenarcic B., Ritonja A., Strukelj B., Turk B., Turk V.,				
RT	J. Biol. Chem. 273:12682-12682(1998).				
RL	-I- FUNCTION: POTENT INHIBITOR OF PAIN-LIKE CYSTEINE PROTEINASES				
CC	AND HAS A PT OF 4.7.				
CC	-I- SIMILARITY: CONTAINS 3 THYROGLOBULIN TYPE-I DOMAINS.				
CC	InterPro: IPR000716; Thyroglobulin_1.				
DR	PIfam: PF00086; thyroglobulin_1; 3.				
DR	SMART: SM00211; TY; 3.				
DR	PROSITE: PS00484; THYROGLOBULIN_1; 2.				
KW	Thiol protease inhibitor; Repeat.				
FT	DOMAIN 18 66 THYROGLOBULIN TYPE 1 1.				
FT	DOMAIN 86 134 THYROGLOBULIN TYPE 1 2.				
FT	DOMAIN 155 199 THYROGLOBULIN TYPE 1 3.				
FT	VARIANT 3 3 S -> T.				
FT	VARIANT 42 42 S -> V.				
FT	VARIANT 130 130 S -> T.				
FT	VARIANT 137 137 P -> E.				
FT	VARIANT 143 144 LQ -> IK.				
FT	VARIANT 156 156 V -> L.				
FT	VARIANT 184 184 G -> D.				
SO	SEQUENCE 199 AA; 21755 MW; 87E44735F8171BRC CRC64;				

Query Match 84.6%; Score 1089; DB 1; Length 199;
Best Local Similarity 95.5%; Pred. No. 2.76-90;
Matches 190; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY	33	SUTKCOQLQASNSGLIGTYVPOCKETEFEEFKOCWSTGYCWCVDDEGKELGTGKINS	92
DB	1	SLSKCOQLQASNSGLIGTYVPOCKETEFEEFKOCWSTGYCWCVDDEGKELGTGKINS	60
QY	93	PCSSRRKALTLTCOMMGAITIVVPWCGPPSCADGSDFEVQCCASNGECYCVDRKREL	152
DB	61	PCSSRRKALTLTCOMMGAITIVVPWCGPPSCADGSDFEVQCCASNGECYCVDRKREL	120

```

OY 153 EGTRGCGPTCEHRLSECEARLKAHNSLREVMPECELDGSGYNPQCMPSGCMCV 212
DB 121 EGTROKGRSPCEHRLSPCEARLKAHNSLREVMPECELDGSGYNPQCMPSGCMCV 180
OY 213 DEGGVAVPGSDVAREKRPCTC 231
DB 181 DEGGVAVPGSDVAREKRPCTC 199

RESULT 2
THYG_MOUSE STANDARD: PR: 2766 AA.
ID THYG_MOUSE 008710; 008590; 090WY7;
AC 01-NOV-1997 (Rel. 35, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYROGLOBULIN PRECURSOR.
GN TG OR TGN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98008184; PubMed=9344706;
RA Caturegli P., Vidalain P.O., Valli M., Aguilera-Galaviz L.A.,
RA Rose N.R.,
RT Cloning and characterization of murine thyroglobulin cDNA.
RL Clin. Immunol. Immunopathol. 85:221-226(1997).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT GOTTER PRO-2283.
RC STRAIN-COG: TISSUE=Thyroid;
RX MEDLINE=98374279; PubMed=9707574;
RA Kim P.S., Hossain S.A., Park Y.-N., Lee I., Yoo S.-E., Arvan P.;
RT thyroglobulin causes congenital goiter with hypothyroidism in the
RT cog/cog mouse: a model of human endoplasmic reticulum storage
RT diseases.
RL Proc. Natl. Acad. Sci. U.S.A. 95:9909-9913(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-AKR/J; TISSUE=Thyroid;
RA Hossain S.A., Yoo S.-E., Kim P.S.;
RT Cloning, characterization, site-directed mutagenesis, and transient
RT expression of 8301-nucleotide AKR/J mouse thyroglobulin cDNA:
RT defective secretion of mutant thyroglobulins.
RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PRECURSOR OF THE IODINATED THYROID HORMONES THYROXINE
CC (T4) AND TRIIODOTHYRONINE (T3).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: THYROID GLAND SPECIFIC (BY SIMILARITY).
CC -1- PTM: SULFATED (BY SIMILARITY).
CC -1- DISEASE: SOME FORMS OF GOTTER, AN ENLARGEMENT OF THE THYROID
CC GLAND, ARE PRODUCED BY DEFECTS IN TG. THE VARIANT P-2284 EXHIBIT A
CC DEFECT IN EXIT FROM THE ENDOPLASMIC RETICULUM.
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYL-ESTERASE/LIPASE FAMILY.
CC -1- SIMILARITY: CONTAINS 11 THYROGLOBULIN TYPE-I DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U76389; AAB53204.1; -
DR EMBL: AF076186; AAC32268.1; -
DR EMBL: AF076187; AAC32269.1; -
DR HSSP: P21836; IMAH.
DR MGD: MGI:98733; Tgn.
DR InterPro: IPR002018; Carboxylesterase_B.

```

```

DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF00135; Coesterase_2.
DR Pfam: PF00086; thyroglobulin_1; 8.
DR SMART: SM00211; TY; 10.
DR PROSITE: PS00484; THYROGLOBULIN_1; 9.
DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.
KW Glycoprotein; Repeat; Thyroid hormone; Iodination; Sulfation; Signal;
KW Disease mutation.
FT SIGNAL 1 20
FT CHAIN 21 2766
FT DOMAIN 32 93
FT DOMAIN 94 161
FT DOMAIN 162 298
FT DOMAIN 299 359
FT DOMAIN 605 658
FT DOMAIN 659 726
FT DOMAIN 727 922
FT DOMAIN 923 1074
FT DOMAIN 1075 1146
FT DOMAIN 1147 1211
FT REPEAT 1455 1468
FT REPEAT 1469 1485
FT REPEAT 1486 1502
FT REPEAT 1511 1564
FT REPEAT 1602 1722
FT REPEAT 1723 1889
FT REPEAT 1890 1992
FT REPEAT 1993 2125
FT REPEAT 2126 2183
FT DISULFID 2263 2280
FT MOD_RES 25 25
FT MOD_RES 2572 2572
FT MOD_RES 2586 2586
FT MOD_RES 2764 2764
FT CARBOHYD 111 111
FT CARBOHYD 139 139
FT CARBOHYD 199 199
FT CARBOHYD 484 484
FT CARBOHYD 496 496
FT CARBOHYD 748 748
FT CARBOHYD 817 817
FT CARBOHYD 948 948
FT CARBOHYD 1141 1141
FT CARBOHYD 1349 1349
FT CARBOHYD 1365 1365
FT CARBOHYD 1715 1715
FT CARBOHYD 1729 1729
FT CARBOHYD 1773 1773
FT CARBOHYD 1864 1864
FT CARBOHYD 1935 1935
FT CARBOHYD 2010 2010
FT CARBOHYD 2120 2120
FT CARBOHYD 2249 2249
FT CARBOHYD 2294 2294
FT CARBOHYD 2581 2581
FT VARIANT 2283 2283
FT CONFLICT 80 80
FT CONFLICT 92 92
FT CONFLICT 1327 1327
FT CONFLICT 1427 1427
FT CONFLICT 1442 1442
FT CONFLICT 1721 1721
FT CONFLICT 1813 1813
FT CONFLICT 1957 1959
FT CONFLICT 2090 2090
FT CONFLICT 2407 2407
FT CONFLICT 2414 2414
FT CONFLICT 2427 2427
FT CONFLICT 2434 2434
FT CONFLICT 2443 2443
FT CONFLICT 2453 2453
FT CONFLICT 2728 2728
FT CONFLICT 2766 2766
SQ SEQUENCE 2766 AA; 304512 MW; E195351DC5793M03 CRC64;
D -> GN (IN REF. 1).
MW: E195351DC5793M03 CRC64;

```

```

Query Match = 23.0%; Score 295.5; DB 1; Length 766;
Best Local Similarity 30.3%; Pred No. 2e-18;
Matches 76; Conservative 31; Mismatches 83; Indels 61; Gaps

QY      16 VMTVLEFACAITTSALTRCQ-----DLQASANSGLIGTVPOCKETGFEERKQ 66
       ::::|:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
Db      4 LVMVSTLLSSVCLVAAMIFEEYOADAOLPRLCELOREKAFLKQAEVFYPCSDSGSFGTQV 63
QY      67 CMCSTGYCCVCVDDEKKEILGIKINGSPPDCSRRAALLITCOM-NQALI---VNVGWCGCP 124
       + ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      64 QANDGSQCVCVSDRREPGRQLGRP-----IYCISFCOLHKORILLASTINSTDALYL 118
QY      122 PSCKADGSEVOCCASNGECYVKDKKGKELEGTRQGRPT-CERHLSCE-EARIKAS 175
       +-+::+::+::+::+::+::+::+::+::+::+::+::+::+::+::+::+::+::+::+::
Db      119 POCODSGNAPAYQCDLQHVQCVCVDTEGMEYVGNROOGRPTRCPR---SCAIRNRLLHG 177
QY      180 NSLRVEMFPEELCEEGSNPVQC-----MPS- 205
       +::+::+::+::+::+::+::+::+::+::+::+::+::+::+::+::+::+::+::+::
Db      176 VGRSP---PQTADGEFMFPVQCFKFNVTIDMALFDLIHNFRPDFAFYTFSSFRGRPEV 233
QY      206 TGYCVWCVEGG 216
       :|||::|::|
Db      233 SGYCYCADSQG 243

RESULT 3
THYG_BOVIN STANDARD: PRT; 2769 AA.
AC PO1267; 018976; Q95478; Q28186;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 02, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYROGLOBULIN PRECURSOR.
TG
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85296288; Pubmed=3855243;
RA Mercken L., Simons M.-J., Swillens S., Massaeer M., Vassart G.;
RT "Primary structure of bovine thyroglobulin deduced from the sequence
   of its 8,431-base complementary DNA.";
RL Nature 316:647-651(1985).
[2]
RP SEQUENCE OF 1-930 FROM N.A.
RX MEDLINE=85127025; Pubmed=3855750;
RA Mercken L., Simons M.-J., de Martynhoff G., Swillens S., Vassart G.;
RT "Presence of hormonogenic and repetitive domains in the first 930
   amino acids of bovine thyroglobulin as deduced from the cDNA
   sequence.";
RL Eur. J. Biochem. 147:59-64(1985).
[3]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE=87190432; Pubmed=3036224;
RA de Martynhoff G., Pohl V., Mercken L., van Ommen G.-J., Vassart G.;
RT "Structural organization of the bovine thyroglobulin gene and of its
   5'-flanking region";
RL Eur. J. Biochem. 164:591-599(1987).
[4]
RP SEQUENCE OF 1002-1209 FROM N.A.
RX MEDLINE=88062712; Pubmed=3681978;
RA Parma J., Christophe D., Pohl V., Vassart G.;
RT "Structural organization of the 5' region of the thyroglobulin gene.
   Evidence for intron loss and 'exonization' during evolution.";
RT J. Mol. Biol. 196:769-779(1987).
CC -1- FUNCTION: PRECURSOR OF THE IODINATED THYROID HORMONES THYROXINE
   (T4) AND TRIIODOTHYRONINE (T3).
CC -1- SUBUNIT: HOMODIMER.
```

	-	SUBCELLULAR LOCATION:	SECRETED.	
CC	-	TISSUE SPECIFICITY:	THYROID GLAND SPECIFIC.	
CC	-	PTM:	SULFATED (BY SIMILARITY).	
CC	-	SIMILARITY:	BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPAS FAMILI	
CC	-	SIMILARITY:	CONTAINS 11 THYROGLOBULIN TYPE-1 DOMAINS.	
CC	-	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb.sib.ch).		
DR	EMBL;	X02815;	CNA26594.1; .-	
DR	EMBL;	X02155;	CNA26090.1; .-	
DR	EMBL;	X05380;	CNA28971.1; ALT_SEQ.	
DR	EMBL;	X06071;	CNA29457.1; .-	
DR	EMBL;	X06072;	CNA29457.1; JOINED.	
DR	EMBL;	X06073;	CNA29457.1; JOINED.	
DR	EMBL;	X06075;	CNA29457.1; JOINED.	
DR	PIR;	A01533;	OJDO.	
DR	HSSP;	P21836;	IMAH.	
DR	InterPro;	IPR002018;	Carboxylesterase_B.	
DR	pfam;	PF00135;	Coesterase_2.	
DR	pfam;	PF00086;	thyroglobulin_1; 8.	
DR	SMART;	SM00211;	TY; 10.	
DR	PROSITE;	PS00484;	THYROGLOBULIN_1; 9.	
RW	GlycoProtex;	PS00941;	CARBHOXYLESTERASE_B_2; 1.	
KM	Glycoprotein Repeat;	Thyroid hormone Iodination;	Sulfation; Signal.	
FT	CHAIN	1..19	THYROGLOBULIN.	
FT	DOMAIN	31..92	THYROGLOBULIN TYPE IA 1.	
FT	DOMAIN	93..160	THYROGLOBULIN TYPE IA 2.	
FT	DOMAIN	161..297	THYROGLOBULIN TYPE IA 3.	
FT	DOMAIN	298..358	THYROGLOBULIN TYPE IA 4.	
FT	DOMAIN	604..657	THYROGLOBULIN TYPE IA 5.	
FT	DOMAIN	661..725	THYROGLOBULIN TYPE IA 6.	
FT	DOMAIN	726..921	THYROGLOBULIN TYPE IA 7.	
FT	DOMAIN	922..1073	THYROGLOBULIN TYPE IA 8.	
FT	DOMAIN	1074..1145	THYROGLOBULIN TYPE IB 1.	
FT	DOMAIN	1146..1210	TYPE II.	
FT	REPEAT	1458..1471	TYPE II.	
FT	REPEAT	1472..1488	TYPE II.	
FT	REPEAT	1489..1505	TYPE II.	
FT	DOMAIN	1513..1567	THYROGLOBULIN TYPE IB 2.	
FT	REPEAT	1605..1725	TYPE IIIA.	
FT	REPEAT	1726..1893	TYPE IIIB.	
FT	REPEAT	1894..1996	TYPE IIIA.	
FT	REPEAT	1997..2130	TYPE IIIB.	
FT	REPEAT	2131..2188	TYPE IIIA.	
FT	MOD_RES	24..24	IODINATION (IN T4).	
FT	MOD_RES	24..24	SULFATION (POTENTIAL).	
FT	MOD_RES	2574..2574	IODINATION (IN T4).	
FT	MOD_RES	2588..2588	IODINATION (IN T4).	
FT	MOD_RES	2767..2767	IODINATION (IN T3).	
FT	CARBOHYD	110..110	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	198..198	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	483..483	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	495..495	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	747..747	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	853..853	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	947..947	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	1140..1140	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	1365..1365	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	1776..1776	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	1870..1870	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	2014..2014	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	2123..2123	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	2251..2251	N-LINKED (GLCNAC...)	(POTENTIAL) .
FT	CARBOHYD	2296..2296	N-LINKED (GLCNAC...)	(POTENTIAL) .


```

FT CARBOHYD 1141 1141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1349 1349 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1365 1365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1715 1715 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1773 1773 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1866 1866 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1937 1937 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2012 2012 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2122 2122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2251 2251 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2296 2296 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2445 2445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2583 2583 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 44 44 L -> V (IN REF. 4).
FT CONFLICT 678 678 A -> V (IN REF. 2).
FT CONFLICT 1492 1492 C -> F (IN REF. 2).
FT CONFLICT 1732 1732 CC -> KS (IN REF. 2).
FT CONFLICT 1914 1914 L -> F (IN REF. 5).
FT CONFLICT 2043 2043 R -> A (IN REF. 5).
FT CONFLICT 2081 2081 Q -> K (IN REF. 5).
FT CONFLICT 2126 2126 A -> V (IN REF. 5).
FT CONFLICT 2153 2153 R -> K (IN REF. 5).
FT CONFLICT 2169 2169 S -> N (IN REF. 5).
FT CONFLICT 2611 2611 M -> I (IN REF. 5).
FT CONFLICT 2658 2658 Q -> H (IN REF. 5).
SQ SEQUENCE 2768 AA; 304642 MW; 290DD6943F3F3D CRC64;

Query Match 22.9%; Score 294.5; DB 1; Length 2768;
Best Local Similarity 30.3%; Pred. No. 2,4e-18;
Matches 76; Conservative 32; Mismatches 82; Indels 61; Gaps 10;

QY 16 VAIWVLFICAITTSTASITKQ-----OLQASANSGLIGTYPPCKEFGEEFKO 66
DB 4 LVIVSTLLSVCLVIANIFEVQVDAQPLRCELQREKAFKODEVFPQCSSEGSFQTVQ 63
QY 67 CWGSTGYCWCVDGDKELGTRKSPDCSRKAALTLCOM-QQAI--VVPWPCGP 121
DB 64 CONDGSCWCVDGDEGVPGRQLGRP-----TACLSFCQLHQRILLSSYSTDALVL 118
QY 122 PSCKADGSDFEVQCCASNGECYCDKKELEGTROGGRPT-CEPHLSDEE-EARKKALS 179
DB 119 POCODSGNAPVQCDLQOVQWCVDTESEMEVYGTROGGRPTCRP---SCEIKSRRLHGG 175
QY 180 NSLRVEMFVPECLIEDGSGYNPVQC-----WPS- 205
DB 176 VG---DKSPQCQDADGEFMPVQCKFVNTDMIFDLIHYNRFPDAVFPSAFNRFPV 232
QY 206 TGYCWCVDGEG 216
DB 233 SGICYCQADSG 243

RESULT 5
THYG_HUMAN STANDARD: PRT; 2768 AA.
AC P01266: O15593; O43899; O15274; Q9NRY1; Q9NRY2; Q9UWZ0; Q9UNY3;
DT 21-JUL-1986 (Rel. 01, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE THYROGLOBULIN PRECURSOR.
GN TG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8724630; PubMed=359559;
RA Malthery Y., Lissitzky S.;
RT "Primary structure of human thyroglobulin deduced from the sequence
RT of its 8448-base complementary DNA.";
RL Eur. J. Biochem. 165:491-498(1987).

```

```

RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=97329813; PubMed=9186272;
RA van de Graaf S.A.R., Pauw E., de Vijlder J.J.M., Ris-Stalpers C.;
RT "The revised 8307 base pair coding sequence of human thyroglobulin
RT transiently expressed in eukaryotic cells.";
RL Eur. J. Endocrinol. 136:508-515(1997).
RN [3]
RN SEQUENCE FROM N.A.
RA Blechschmidt K., Kalaydjieva L., Goodman R., Gresham D., Baas F.,
RA de Jonge R., Schlijhabel M.B., Menzel U., Dettle M.D., Baumgart C.,
RA Jahn N., Rosenthal A.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE OF 1-730 FROM N.A.
RX MEDLINE=85127024; PubMed=3971976;
RA Malthery Y., Lissitzky S.;
RT "Sequence of the 5'-end quarter of the human-thyroglobulin messenger
RT ribonucleic acid and of its deduced amino-acid sequence.";
RL Eur. J. Biochem. 147:53-58(1985).
RN [5]
RN SEQUENCE OF 1-415; 640-737 AND 880-1000 FROM N.A.
RX MEDLINE=88062712; PubMed=3681978;
RA Parma J., Christophe D., Pohl V., Vassart G.;
RT "Structural organization of the 5' region of the thyroglobulin gene.
RT Evidence for intron loss and 'exonization' during evolution.";
RL J. Mol. Biol. 196:769-779(1987).
RN [6]
RN SEQUENCE OF 1-25 FROM N.A.
RX MEDLINE=85269632; PubMed=2991855;
RA Christophe D., Cabrer B., Baccolle A., Targovnik H.M., Pohl V.,
RA Vassart G.;
RT "An unusually long poly(purine)-poly(pyrimidine) sequence is located
RT upstream from the human thyroglobulin gene.";
RL Nucleic Acids Res. 13:5127-5144(1985).
RN [7]
RN SEQUENCE OF 1002-1566 FROM N.A.
RX PubMed=11124863;
RA Moya C.M., Mendive F.M., Rivolta C.M., Vassart G., Targovnik H.M.;
RT "Genomic organization of the 5' region of the human thyroglobulin
RT gene.";
RL Eur. J. Endocrinol. 143:789-798(2000).
RN [8]
RN SEQUENCE OF 1645-2768 FROM N.A.
RX MEDLINE=99452200; PubMed=10524569;
RA Mendive F.M., Rivolta C.M., Vassart G., Targovnik H.M.;
RT "Genomic organization of the 3' region of the human thyroglobulin
RT gene.";
RL Thyroid 9:903-912(1999).
RN [9]
RN PRESENCE OF A 11TH TYROGLOBULIN TYPE-I REPEAT.
RX MEDLINE=96390872; PubMed=8797845;
RA Molina F., Bouanani M., Pau B., Granier C.;
RT "Characterization of the type-1 repeat from thyroglobulin, a
RT cysteine-rich module found in proteins from different families.";
RL Eur. J. Biochem. 240:125-133(1996).
RN [10]
RN IODINATION SITES.
RX MEDLINE=89340430; PubMed=2760035;
RA Lamas L., Anderson P.C., Fox J.W., Dunn J.T.;
RT "Consensus sequences for early iodination and hormonogenesis in human
RT thyroglobulin.";
RL J. Biol. Chem. 264:13541-13545(1989).
RN [11]
RN SULFATION.
RX PubMed=10448091;
RA Nlend M.-C., Cauvi D., Venot N., Chabaud O.;
RT "Sulfated tyrosines of thyroglobulin are involved in thyroid hormone
RT synthesis.";
RL Biochem. Biophys. Res. Commun. 262:193-197(1999).
RN [12]
RN VARIANT GOITER HIS-870.

```

RX MEDLINE-93164775; PubMed-8094490;
 RA Corral J., Martin C., Perez R., Gonzalez-Santiago R.;
 RA San Millan J.L., Miralles J.M., Gonzalez-Santiago R.;
 RT "Thyroglobulin gene point mutation associated with non-endemic simple
 RT goiter";
 RL Lancet 341:462-464(1993).
 RN (13)
 RP VARIANTS: GOTTER R-1264 AND S-1996, AND VARIANTS.
 RX MEDLINE-9213884; PubMed-10199792;
 RA Hishinuma A., Takametsu J., Ohshima Y., Yokozawa T., Kanno Y., Kuma K.,
 RA Yoshida S., Matsura N., Ielri T.;
 RT "Two novel cysteine substitutions (C1263R and C1995S) of thyroglobulin
 RT cause a defect in intracellular transport of thyroglobulin in
 RT patients with congenital goiter and the variant type of adenomatous
 RT goiter";
 RL J. Clin. Endocrinol. Metab. 84:1438-1444(1999).
 CC -1- FUNCTION: PRECURSOR OF THE IODINATED THYROID HORMONES THYROXINE
 CC (T4) AND TRIIODOTHYRONINE (T3).
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: THYROID GLAND SPECIFIC.
 CC -1- PTM: SULFATED.
 CC -1- DISPAR: DEFECTS IN ITS CAUSE OF FORMS OF GOTTER, AN
 CC ENLARGEMENT OF THE THYROID GLAND. THIS IS SOMETIMES LINKED WITH
 CC HYPOTHYROIDISM.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 11 THYROGLOBULIN TYPE-I DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X05615; CAA29104.1; -;
 DR EMBL: U93033; AAC51924.1; -;
 DR EMBL: AF230667; AAF62513.2; -;
 DR EMBL: AF231007; AAG29877.1; -;
 DR EMBL: AF230666; AAF62515.1; -;
 DR EMBL: AF305872; AAG17932.1; -;
 DR EMBL: X02154; CAA25089.1; -;
 DR EMBL: X06059; CAA29454.1; -;
 DR EMBL: X06060; CAA29454.1; JOINED.
 DR EMBL: X06061; CAA29454.1; JOINED.
 DR EMBL: X06062; CAA29454.1; JOINED.
 DR EMBL: X06063; CAA29454.1; JOINED.
 DR EMBL: X06064; CAA29454.1; JOINED.
 DR EMBL: X06065; CAA29454.1; JOINED.
 DR EMBL: X06066; CAA29454.1; JOINED.
 DR EMBL: X06067; CAA29455.1; -;
 DR EMBL: X06068; CAA29455.1; JOINED.
 DR EMBL: X06069; CAA29455.1; JOINED.
 DR EMBL: X06070; CAA29456.1; JOINED.
 DR EMBL: X02749; CAA26527.1; -;
 DR EMBL: X02749; CAA26527.1; -;
 DR EMBL: AF170489; AAD51647.1; -;
 DR EMBL: AF170486; AAD51647.1; JOINED.
 DR EMBL: AF170487; AAD51647.1; JOINED.
 DR EMBL: AF170488; AAD51647.1; JOINED.
 DR EMBL: AF105687; AAC95473.1; -;
 DR EMBL: AF105681; AAC95473.1; JOINED.
 DR EMBL: AF105682; AAC95473.1; JOINED.
 DR EMBL: AF105683; AAC95473.1; JOINED.
 DR EMBL: AF105684; AAC95473.1; JOINED.
 DR EMBL: AF105685; AAC95473.1; JOINED.
 DR EMBL: AF105686; AAC95473.1; JOINED.
 DR EMBL: AF080484; AAD50912.2; JOINED.
 DR EMBL: AF169654; AAD50912.2; JOINED.
 DR EMBL: AF169655; AAD50912.2; JOINED.
 DR EMBL: AF169656; AAD50912.2; JOINED.
 DR EMBL: AF169657; AAD50912.2; JOINED.
 DR EMBL: AF169658; AAD50912.2; JOINED.

DR EMBL: AF169659; AAD50912.2; JOINED.
 DR EMBL: AF169661; AAD50912.2; JOINED.
 DR EMBL: AF169662; AAD50912.2; JOINED.
 DR EMBL: AF169663; AAD50912.2; JOINED.
 DR EMBL: AF169664; AAD50912.2; JOINED.
 DR EMBL: AF080472; AAD50912.2; JOINED.
 DR EMBL: AF080473; AAD50912.2; JOINED.
 DR EMBL: AF080474; AAD50912.2; JOINED.
 DR EMBL: AF080475; AAD50912.2; JOINED.
 DR EMBL: AF080476; AAD50912.2; JOINED.
 DR EMBL: AF080477; AAD50912.2; JOINED.
 DR EMBL: AF080478; AAD50912.2; JOINED.
 DR EMBL: AF080479; AAD50912.2; JOINED.
 DR EMBL: AF080480; AAD50912.2; JOINED.
 DR EMBL: AF080481; AAD50912.2; JOINED.
 DR EMBL: AF080482; AAD50912.2; JOINED.
 DR EMBL: AF080483; AAD50912.2; JOINED.
 DR PIR: A01532; UTH0.
 DR PIR: S00014; S00014.
 DR PIR: S03422; S03422.
 DR HSSP: P21836; 1MAH.
 DR MIM: 188450; -;
 DR MIM: 138800; -;
 DR MIM: 274900; -;
 DR InterPro: IPR002018; Carboxylesterase-B.
 DR InterPro: IPR00716; Thyroglobulin_1.
 DR Pfam: PF00135; Coesterase_2.
 DR Pfam: PF00086; thyroglobulin_1; 8.
 DR SMART: SM00211; TY; 10.
 DR PROSITE: PS00484; THYROGLOBULIN_1; 9.
 DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.
 KW Glycoprotein; Repeat; Thyroid hormone; Iodination; Sulfation; Signal;
 KW Disease mutation; Polymorphism.
 FT SIGNAL 1 19
 FT CHAIN 20 2768 THYROGLOBULIN.
 FT DOMAIN 32 92 THYROGLOBULIN TYPE 1A 1.
 Query Match 21.1%; Score 272; DB 1; Length 2768;
 Best Local Similarity 26.7%; Pred. No. 2.5e-16;
 Matches 75; Conservative 32; Mismatches 80; Indels 94; Gaps 9;
 QY 32 ASITKQ-OLQSANSGIG---TYVPOCKETGEFEKOCWGSYGCWCDDEKREILG 86
 Db 91 ACISFQLOKQOILSLSTINSTDISTYIPQCGDGAAPQCVQYQVQCMCDAGMAYVG 150
 QY 87 TKIRGSPD-CRRKALTLCCOMQALIVNPGMCGPPSCADSPDEYQCCASN----- 139
 Db 151 TROLGPRKPCRS-----CEIRNRRLHGVGRKSPQCSAGFEMPVQCKFVNTDMT 204
 QY 140 -----GECYCVDKKGELEGTROQ----- 158
 Db 205 FDLVHSNFPDPAFVTFSSPGRREPEVSGYCHCADSGRELAETGLLELDIYDTIFAG 264
 QY 159 -----GPPCEIRLSPCEARIRKANSLSAREMVEPE 190
 Db 265 LIDPSTFTTTLIRLQRRFLANOVSYSGFRCP---TKCEKERTATISFG---HPVPS 318
 QY 191 CLEDSYNEVQCPSTGCMVCGGVKYPGSDVRRKRPIC 231
 Db 319 CRNMGDYQAVQC-QTEGSCWCVDAQGEMHGTROQDEPPSC 358
 RESULT 6
 NID2_HUMAN
 ID NID2_HUMAN STANDARD; PRT; 1375 AA.
 AC Q14112: 043710;
 DT 15-DEC-1998 (rel. 37, Created)
 DT 20-APR-2001 (rel. 40, Last sequence update)
 DT 20-APR-2001 (rel. 40, Last annotation update)
 DE NIDOGEN-2 PRECURSOR (NID-2) (OSTEONIDOGEN).
 GN NID2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

DB 1054 CHECKSDPCWCDKDEVOGFT 1074

RESULT 7
SAX_RANCA STANDARD: PRT: 844 AA.
AC P31226:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SAXIPHILIN PRECURSOR (SAX).
OC Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94195765; PubMed=8146142;
RA Morabito M.A., Moczydlowski E.;
RT "Molecular cloning of bullfrog saxiphilin: a unique relative of the
transferrin family that binds saxitoxin.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:2478-2482(1994).
RN [2]
RP REVISIONS:
RX MEDLINE=95327702; PubMed=7604048;
RA Morabito M.A., Moczydlowski E.;
RL Proc. Natl. Acad. Sci. U.S.A. 92:6651-6651(1995).
RN [3]
RP SEQUENCE OF 20-44; 317-333; 360-371; 541-571; 606-624 AND 690-714.
RC TISSUE=Plasma;
RX MEDLINE=91332078; PubMed=1869567;
RA Li Y., Moczydlowski E.;
RT "Purification and partial sequencing of saxiphilin, a
saxitoxin-binding protein from the bullfrog, reveals homology to
transferrin.";
RL J. Biol. Chem. 266:15481-15487(1991).
CC -1- FUNCTION: BINDS SPECIFICALLY TO THE NEUROTOXIN SAXITOXIN. ITS
PHYSIOLOGICAL ROLE MAY BE TO TRANSPORT OR SEQUESTER AN ENDOGENOUS
ORGANIC MOLECULE OTHER THAN FE(3+). IT MAY PARTICIPATE IN A
DETOXIFICATION MECHANISM FOR NEUTRALIZING A MICROBIAL TOXIN.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: PLASMA. HIGHEST LEVELS OF TRANSCRIPTS FOUND IN
THE LIVER, THE LUNG, THE PANCREAS AND THE BRAIN.
CC -1- SIMILARITY: CONTAINS 2 THYROGLOBULIN TYPE-1 DOMAINS.
CC -1- SIMILARITY: RELATED TO TRANSFERRIN (BUT DOES NOT BIND FE(3+)).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch)
CC -----
DR EMBL: U05246; AAA75440.1;
DR PIR: A39426; A39426.
DR HSP: P56410; IAOV.
DR InterPro: IPR000716; Thyroglobulin_1.
DR InterPro: IPR001156; Transferrin.
DR Pfam: PF00086; thyroglobulin_1; 2.
DR Pfam: PF00405; transferrin; 4.
DR PRINTS: PR00422; TRANSFERRIN.
DR SMART: SM00094; TR_FER. 2.
DR SMART: SM00211; TY_2.
DR PROSITE: PS00484; THYROGLOBULIN_1; 2.
KW Signal; Extracellular matrix; Repeat.
FT CHAIN 1 19 SAXIPHILIN.
FT CHAIN 20 844 THYROGLOBULIN TYPE I 1.
FT DOMAIN 124 172 THYROGLOBULIN TYPE I 2.
FT DOMAIN 196 244 THYROGLOBULIN TYPE I 2.

FT DOMAIN 109 251 ABSENT IN TRANSFERRINS.
FT DISULFID 29 64 BY SIMILARITY.
FT DISULFID 39 55 BY SIMILARITY.
FT DISULFID 277 360 BY SIMILARITY.
FT DISULFID 322 335 BY SIMILARITY.
FT DISULFID 332 343 BY SIMILARITY.
FT DISULFID 388 402 BY SIMILARITY.
FT DISULFID 495 527 BY SIMILARITY.
FT DISULFID 505 518 BY SIMILARITY.
FT DISULFID 552 839 BY SIMILARITY.
FT DISULFID 570 799 BY SIMILARITY.
FT DISULFID 607 685 BY SIMILARITY.
FT DISULFID 641 665 BY SIMILARITY.
FT DISULFID 652 668 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
SQ SEQUENCE 844 AA; 93089 MW; 162A765AD02C3C5E CRC64;

Query Match 19.0%; Score 244; DB 1; Length 844;
Best Local Similarity 28.9%; Pred. No. 2,4e-14;
Matches 57; Conservative 25; Mismatches 47; Indels 68; Gaps 5;

QY 29 STEASLTKC-QOLASANSGLGTYVPOCKEETGEFEERQCGSTGYCWCYDEDEGKEIIGT 87
| : | | : : | : | | | | : : | : | | | | : : | : | | | | : : | : | |
DB 102 SSNDLQKCKLKEHQALAKMIGHIYIPQCDKGNQPOQCGSGTGHCMVNAAGEKISGT 161
QY 88 KIRSPDCSRKALTLCOMQAIIVNPGWCSPCKADGSEDEVOCASNGCYCYVDK 147
DB 162 NT-----PG-----166
QY 148 KGELEGTROGPRGPRH-TSEC-LEARIKANSLSLVEMFVEPECLEDSYNVPCWPS 205
| : | | | | : : | : | : : : | : | : : | : | : : | : | : : | : | :
DB 167 -----QATRCERHDLPCCKERGVAGDEKVIAGFVQCDKGNVPEQGFHS 216
QY 206 TGVCWCYDEGVKVPWS 222
DB 217 TGYSVCVNAIGEETAG 233

RESULT 8
NID2_MOUSE STANDARD: PRT: 1403 AA.
AC NID2_MOUSE
ID O88322;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NID2_MOUSE (NID-2) (ENTACTIN-2).
GN NID2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98297167; PubMed=9633511;
RA Kimura N., Toyoshima T., Kojima T., Shimane M.;
RT "Entactin-2, a new member of basement membrane protein with high
homology to entactin/nidogen.";
RL Exp. Cell Res. 241:36-45(1998).
CC -1- FUNCTION: CELL ADHESION GLYCOPROTEIN. MIGHT BE INVOLVED IN
OSTEOBLAST DIFFERENTIATION. IT PROBABLY HAS A ROLE IN CELL-
EXTRACELLULAR MATRIX INTERACTIONS.
CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANE.
CC -1- PTM: HIGHLY N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 THYROGLOBULIN TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

[illegible]

```

Db 1038 -----EFT-GRPTVEER-----WRSLHEHGGTFRPDQYVPCDDLGHTIFPLQJCKGSD 1087
Oy 208 YCMCVDEGGVYKVPGSVDR-FKRPCT 231
Db 1088 FCMCVCDKDGRELQGTFRSGPTFRPAC 1112

RESULT 9
HG2A_MOUSE STANDARD: PRT, 279 AA.
ID HG2A_MOUSE
AC P04441;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN, GAMMA CHAIN (MHC CLASS II(A)
DE ASSOCIATED INVARIANT CHAIN).
OS CD74 OR II.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Sciurognathi; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AKR; TISSUE=Liver;
RC MEDLINE=89098406; PubMed=2492095;
RA Zhu L., Jones P.P.;
RL "Complete sequence of the murine invariant chain (II) gene.";
RL Nucleic Acids Res. 17:447-448(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AKR;
RC MEDLINE=87275861; PubMed=3038530;
RA Koch N., Laner W., Habicht J., Dobberstein B.;
RA "Primary structure of the gene for the murine Ia antigen-associated
RA invariant chains (II). An alternatively spliced exon encodes a
RA cysteine-rich domain highly homologous to a repetitive sequence of
RA thyroglobulin.";
RL EMBO J. 6:1677-1683(1987).
RN [3]
RP SEQUENCE OF 1-58 FROM N.A.
RA Stone J., Perry R., Todd J.A., McDevitt H.O.;
RT "Nucleotide sequences of the murine Ia-associated invariant chain (II)
RT and I-E (H-2S, beta) chain expressible cDNA clones.";
RL Submitted (MAR-1988) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE OF 47-191 AND 256-279 FROM N.A.
RA MEDLINE=84207946; PubMed=6327293;
RA Singer P.A., Laner W., Dembic Z., Mayer W.E., Lipp J., Koch N.,
RA Hammerling G., Klein J., Dobberstein B.;
RT "Structure of the murine Ia-associated invariant (II) chain as
RT deduced from a cDNA clone.";
RL EMBO J. 3:873-877(1984).
RN [5]
RP FUNCTION: PLAYS A CRITICAL ROLE IN MHC CLASS II ANTIGEN PROCESSING
RP BY STABILIZING PEPTIDE-FREE CLASS II ALPHA/BETA HETERODIMERS IN A
RP COMPLEX SOON AFTER THEIR SYNTHESIS AND DIRECTING TRANSPORT OF THE
RP COMPLEX FROM THE ENDOPLASMIC RETICULUM TO COMPARTMENTS WHERE
RP PEPTIDE LOADING OF CLASS II TAKES PLACE.
RN [6]
RP SUBUNIT: NONAMER COMPOSED OF THREE ALPHA/BETA/GAMMA HETEROTRIMERS
RP (BY SIMILARITY).
RN [7]
RP SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
RN [8]
RP -1- SIMILARITY: CONTAINS 1 THYOGLOBULIN TYPE-I DOMAIN.
RN [9]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration with
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial
RP entities requires a license agreement (See http://www.isb-sib.ch/announce/
RP or send an email to license@isb-sib.ch).
RN [10]
RP EMBL: X00496; CAA25191.1; -
RP EMBL: X07129; CAA30141.1; -
DR

```

Page 9

DR EMBL: X05428; CAA29010.1; -;
 DR EMBL: X05429; CAA29012.1; -;
 DR EMBL: X05430; CAB37297.1; -;
 DR EMBL: X13414; -; NOT_ANNOTATED_CDS.
 DR PIR: A02244; HLHMSG.
 DR PIR: A27866; A27866.
 DR PIR: B27866; B27866.
 DR PIR: S03099; S03099.
 DR PIR: S09284; S09284.
 DR HSSP: P04233; 1A6A.
 DR MGD: MGI:96534; 11.
 DR Interpro: IPR000716; Thyroglobulin_1.
 DR Pfam: PF00086; thyroglobulin_1; 1.
 DR SMART: SMO0211; TY; 1.
 DR PROSITE: PS00484; THYROGLOBULIN_1; 1.
 DR Glycoprotein: MHC; Proteoglycan; Transmembrane; Alternative splicing;
 KW Signal-anchor.
 FT DOMAIN 1 29 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 30 55 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 56 279 EXTRACELLULAR (POTENTIAL).
 FT CARBOHYD 113 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 265 265 O-LINKED (GLYCOSAMINOGLYCAN).
 FT DOMAIN 211 254 THYROGLOBULIN TYPE I.
 FT VARSPLIC 192 255 MISSING (IN SHORT ISOFORM).
 FT TRANSMEM 10 10 N -> M (IN REF. 1).
 FT CONFLICT 113 113 N -> Q (IN REF. 1).
 FT CONFLICT 229 231 STG -> RHC (IN REF. 2).
 FT CONFLICT 231 231 STG -> RHC (IN REF. 2).
 SQ SEQUENCE 279 AA: 31557 MW: 300485014Fl70792 CRC64:

Query Match 12.1%; Score 156; DB 1; Length 279;
 Best Local Similarity 40.7%; Pred. No. 5,6e-07;
 Matches 35; Conservative 11; Mismatches 32; Indels 8; Gaps 2;

QY 19 WYLFACAT-----STAS---LTKCOQLQASNSGLITTYPOCKEFGEEKQCMG 70
 DB 170 WLFEMSKNSLEKKRPTAPKPVLTCKOEVSHPVAYPGAFPKCDENGNYLPQCHGS 229
 QY 71 TGYCWCVDDEGKEILGIRKSPDSC 96
 DB 230 TGYCWCVFNGTEVPHTRKSRHNC 255

RESULT 10
 NIDO_MOUSE STANDARD; PRT: 1245 AA.
 AC P10493;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NIDOGEN PRECURSOR (ENTACTIN).
 GN NID OR NIDI OR ENT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40.
 RX MEDLINE=89079780; PubMed=3264556;
 RA Durkin M.E., Chakravarti S., Barros B.B., Liu S.H., Friedman R.L.,
 RA Chung A.E.;
 RT "Amino acid sequence and domain structure of entactin. Homology with
 RT epidermal growth factor precursor and low density lipoprotein
 RT receptor.";
 RL J. Cell Biol. 107:2749-2756(1988).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=89231638; PubMed=2496973;
 RA Mann K., Deutzmann R., Altmalley M., Timpl R., Raimondi L.,
 RA Yamada Y., Pan T.-C., Conway D., Chu M.-L.;
 RT "Amino acid sequence of mouse nidogen, a multidomain basement

RT membrane protein with binding activity for laminin, collagen IV and
 RT cells.";
 RL EMBO J. 8:65-72(1989).
 RN [3]
 RP SEQUENCE OF 1-251 FROM N.A.
 RC STRAIN-BALB/C; TISSUE=Liver;
 RX MEDLINE=94040771; PubMed=8224873;
 RA Durkin M.E., Liu S.H., Reing J., Chung A.E.;
 RT "Characterization of the 5' end of the mouse Ent gene encoding the
 RT basement membrane protein, entactin.";
 RL Gene 132:261-266(1993).
 RN [4]
 RP SEQUENCE OF 1207-1245 FROM N.A.
 RC STRAIN=C57BL/6J X CBA/J;
 RX MEDLINE=95324912; PubMed=7601446;
 RA Durkin M.E., Meyer U.M., Chung A.E.;
 RT "Exon organization of the mouse entactin gene corresponds to the
 RT structural domains of the polypeptide and has regional homology to
 RT the low-density lipoprotein receptor gene.";
 RL Genomics 26:219-228(1995).
 RN [5]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=86192477; PubMed=3084254;
 RA Paulsson M., Deutzmann R., Dziadek M., Nowack H., Timpl R., Weber S.,
 RA Engel J.;
 RT "Purification and structural characterization of intact and
 RT fragmented nidogen obtained from a tumor basement membrane.";
 RL Eur. J. Biochem. 156:467-478(1986).
 RN [6]
 RP CARBOHYDRATE-LINKAGE SITES, AND PARTIAL SEQUENCE.
 RX MEDLINE=93316903; PubMed=8326911;
 RA Fujiwara S., Shinkai H., Mann K., Timpl R.;
 RT "Structure and localization of O- and N-linked oligosaccharide chains
 RT on basement membrane protein nidogen.";
 RL Matrix 13:215-222(1993).
 CC -1- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN
 CC BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.
 CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-
 CC EXTRACELLULAR MATRIX INTERACTIONS.
 CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.
 CC -1- PTM: N- AND O-GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR XYND DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC
 CC EMBL: X14194; CAA32408.1; -;
 CC EMBL: X14480; CAA32642.1; -;
 CC EMBL: L17324; AAAT7652.1; -;
 CC EMBL: L17322; AAAT7652.1; JOINED.
 CC EMBL: L17323; AAAT7652.1; JOINED.
 CC EMBL: X83093; CAA58148.1; -;
 CC PIR: S02730; MMSMSND.
 CC HSSP: P07204; 2NDX.
 CC CarDBank; CCSD:37018; -;
 CC CarDBank; CCSD:37019; -;
 CC MGD: MGI:97342; NId1.
 CC Interpro: IPR000152; Asx_hydroxyl.
 CC Interpro: IPR000561; EGF-like.
 CC Interpro: IPR001881; EGF Ca.
 CC Interpro: IPR000033; LDL_receptor_rep.
 CC Interpro: IPR003886; Nidogen_ext.
 CC Interpro: IPR000716; Thyroglobulin_1.
 CC Pfam: PF00008; EGF; 6.
 CC Pfam: PF00058; Idl_recept_b; 3.
 CC Pfam: PF00086; thyroglobulin_1; 1.

Db 771 QRAQCYMGSSYTCSCLPGEFG--DGRARVDVEQGHRCRHPAFCAVNTGSPYCCQKR 828

QY 149 GKRLSTYR---QGGRPLCEHNLSEGEARPAKHSLSLVE-MAPDELDEGSSNPVQC 203

Dd 829 GYGDDGFRCPREVAEKTQCLREHNLILCAAGCADAPRLQGMFVPGDDEGVHVPQCH 888

QY 204 PSTGTGTCYVDEGSAKVPSSDY-REKRPQC 231

Dd 889 HSTGTGTCYVDRGRLSESSKRPQMRPPC 917

RESULT	11			
AC	HG2A-RAT	STANDARD:	PTT:	280 AA.
AC	P10247:			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN, GAMMA CHAIN (HLA-DR ANTIGENS ASSOCIATED INVARIANT CHAIN).			
DE	CD74.			
GN	Rattus norvegicus (Rat).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-PVG X DA; TISSUE=Spleen;			
RX	MEDLINE=89282409; PubMed=2499873;			
RA	McKnight A.J., Mason D.W., Barclay A.N.;			
RT	"Sequence of a rat MHC class II-associated invariant chain cDNA clone containing a 64 amino acid Lhyroglobulin-like domain."			
RL	Nucleic Acids Res. 17:3983-3984(1989).			
RN	[2]			
RP	SEQUENCE OF 1-192 AND 257-280 FROM N.A.			
RC	STRAIN-LEWIS;			
RX	MEDLINE=89098337; PubMed=3264906;			
RA	Henkes W., Syha U., Reske K.;			
R7	"Nucleotide sequence of rat invariant gamma chain cDNA clone pLR gamma 34.3."			
RT	Nucleic Acids Res. 16:11822-11822(1988).			
RL	-1- FUNCTION: PLAYS A CRITICAL ROLE IN MHC CLASS II ANTIGEN PROCESSING			
CC	BY STABILIZING PEPTIDE-FREE CLASS II ALPHA/BETA HETERODIMERS IN A COMPLEX SOON AFTER THEIR SYNTHESIS AND DIRECTING TRANSPORT OF THE COMPLEX FROM THE ENDOPLASMIC RETICULUM TO COMPARTMENTS WHERE PEPTIDE LOADING OF CLASS II TAKES PLACE.			
CC	-1- SUBUNIT: NONAMER COMPOSED OF THREE ALPHA/BETA/GAMMA HETEROTRIMERS (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).			
CC	-1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_licenselsib-sib.ch).			
CC	-----			
DR	EMBL: X14254; CAA32468.1; -			
DR	EMBL: X13044; CAA31450.1; -			
DR	PIR: S02182; S02182.			
DR	PIR: S04362; S04362.			
DR	HSSP: P04233; IAGA.			
DR	Interpro: IPR00716; Thyroglobulin_1.			
DR	Pfam: PF00086; thyroglobulin_1.			
DR	SMART: SM00211; TY; 1.			
DR	PROSITE: PS00484; THYROGLOBULIN_1; 1.			
KW	Glycoprotein; MHC; proteoglycan; Transmembrane; Alternative splicing; Signal anchor.			
FT	DOMAIN 1 30			CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 31 56			SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).

FT DOMAIN 57 280 EXTRACELLULAR (POTENTIAL).
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 266 266 O-LINKED (GLYCOSAMINOGLYCAN) (BY
 SIMILARITY).
 FT DOMAIN 212 255 THYROGLOBULIN TYPE I.
 FT VARSPIC 193 256 MISSING (IN SHORT ISOCORN).
 SO SEQUENCE 280 AA: 31642 MW: D935D169A9B5732 CRC64:

Query Match 11.9%, Score 153, DB 1: Length 280;
 Best Local Similarity 38.4%, Pred. No. 1e-06;
 Matches 33; Conservative 10; Mismatches 35; Indels 8; Gaps 1;

OY 19 WVFIFCAITTEAS-----LTKCOLOASANSGLIGYVPOCKEKEGEPEKQCMGS 70
 DB 171 WLFEMSKNSLEKQPTQTPRVVLTKCOEYSHLPDVHPGAFRCKDENQNYPLQCHS 230
 OY 71 TGYCWCYDEGKEILGTIKRSPDCS 96
 DB 231 TGYCWCYFPNGTEVPHTSRGRHNS 256

RESULT 12
 NIDO HUMAN STANDARD: PRT: 1247 AA.
 AC P14543: Q14942: (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NIDOGEN PRECURSOR (ENTACTIN).
 GN NID.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID:9606;
 RN 11
 RP SEQUENCE FROM N.A. PubMed=2574658;
 RX MEDLINE=90091745; PubMed=2574658;
 RA Nagayoshi T., Sanborn D., Hickok N.J., Olsen D.R., Fazio M.J.,
 Chu M.-L., Knowlton R., Mann K., Deutzmann R., Timpl R., Uitto J.;
 RT "Human nidogen: complete amino acid sequence and structural domains
 deduced from cDNAs, and evidence for polymorphism of the gene.";
 RL DNA 8:581-594(1989).
 RN 12
 RP SEQUENCE FROM N.A. PubMed=7557988;
 RX MEDLINE=96044428; PubMed=7557988;
 RA Zimmermann K., Holschen S., Hainer M., Nischt R.;
 RT "Genomic sequences and structural organization of the human nidogen
 gene (NID).";
 RL Genomics 27:245-250(1995).
 RN 13
 RP SEQUENCE OF 667-1247 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=89270475; PubMed=2471408;
 RA Olsen D.R., Nagayoshi T., Fazio M., Mattei M.-G., Passage E., Weil D.,
 Timpl R., Chu M.L., Uitto J.;
 RT "Human nidogen: cDNA cloning, cellular expression, and mapping of the
 gene to chromosome 14q3.2";
 RL Am. J. Hum. Genet. 44:876-885(1989).
 CC -1- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN
 CC BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.
 CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-
 CC EXTRACELLULAR MATRIX INTERACTIONS.
 CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.
 CC -1- PTM: N- AND O-GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M30269; AAA59932.1; -;
 CC EMBL: X82245; CAAS7709.1; -;
 CC EMBL: X84819; CAAS7709.1; JOINED.
 CC EMBL: X84820; CAAS7709.1; JOINED.
 CC EMBL: X84821; CAAS7709.1; JOINED.
 CC EMBL: X84822; CAAS7709.1; JOINED.
 CC EMBL: X84823; CAAS7709.1; JOINED.
 CC EMBL: X84824; CAAS7709.1; JOINED.
 CC EMBL: X84825; CAAS7709.1; JOINED.
 CC EMBL: X84826; CAAS7709.1; JOINED.
 CC EMBL: X84827; CAAS7709.1; JOINED.
 CC EMBL: X84828; CAAS7709.1; JOINED.
 CC EMBL: X84829; CAAS7709.1; JOINED.
 CC EMBL: X84830; CAAS7709.1; JOINED.
 CC EMBL: X84831; CAAS7709.1; JOINED.
 CC EMBL: X84832; CAAS7709.1; JOINED.
 CC EMBL: X84833; CAAS7709.1; JOINED.
 CC EMBL: X84834; CAAS7709.1; JOINED.
 CC EMBL: X84835; CAAS7709.1; JOINED.
 CC EMBL: X84836; CAAS7709.1; JOINED.
 CC EMBL: X84837; CAAS7709.1; JOINED.
 CC EMBL: M27445; AAA57261.1; -;
 CC PIR: A33322; MMROND.
 CC HSSP: F07204; ZADK.
 CC MIM: 131390; -;
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR001881; EGF Ca.
 CC InterPro: IPR000033; LDL_receptor.
 CC InterPro: IPR003886; Nidogen_ext.
 CC InterPro: IPR00716; Thyroglobulin_1.
 CC Pfam: PF00008; EGF_6.
 CC Pfam: PF00058; ldl_recept_b; 3.
 CC Pfam: PF00086; thyroglobulin_1; 1.
 CC SMART: SM00179; EGF_CA; 2.
 CC SMART: SM00001; EGF_Like; 4.
 CC SMART: SM00135; LY; 5.
 CC SMART: SM00539; NIDO; 1.
 CC SMART: SM00211; TY; 1.
 CC PROSITE: PS00010; ASX_HYDROXYL; 3.
 CC PROSITE: PS00484; THYROGLOBULIN_1; 1.
 CC PROSITE: PS00022; EGF_1; 1.
 CC PROSITE: PS01186; EGF_2; 5.
 CC PROSITE: PS01187; EGF_CA; 2.
 CC KW Basement membrane; Extracellular matrix; Glycoprotein; Sulfation;
 CC KW Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion.
 CC FT SIGNAL 1 28
 CC FT CHAIN 29 1247
 CC FT DOMAIN 29 669 NITROGEN.
 CC FT DOMAIN 670 917 I (LARGER GLOBULAR DOMAIN).
 CC FT DOMAIN 918 1247 IIT (SMALLER GLOBULAR DOMAIN).
 CC FT DOMAIN 386 426 EGF-LIKE 1.
 CC FT DOMAIN 668 709 EGF-LIKE 2.
 CC FT DOMAIN 710 751 EGF-LIKE 3.
 CC FT DOMAIN 752 801 EGF-LIKE 4.
 CC FT DOMAIN 802 840 EGF-LIKE 5.
 CC FT DOMAIN 841 899 THYROGLOBULIN TYPE I.
 CC FT DOMAIN 900 1030 LDL-RECEPTOR YWTD MOTIF 1.
 CC FT DOMAIN 1031 1073 LDL-RECEPTOR YWTD MOTIF 2.
 CC FT DOMAIN 1074 1118 LDL-RECEPTOR YWTD MOTIF 3.
 CC FT DOMAIN 1119 1163 LDL-RECEPTOR YWTD MOTIF 4.
 CC FT DOMAIN 1208 1244 EGF-LIKE 6.
 CC FT MOD_RES 289 289 SULFATION (POTENTIAL).
 CC FT MOD_RES 296 296 SULFATION (POTENTIAL).
 CC FT DISULFID 672 685 BY SIMILARITY.
 CC FT DISULFID 679 695 BY SIMILARITY.
 CC FT DISULFID 708 708 BY SIMILARITY.
 CC FT DISULFID 714 727 BY SIMILARITY.

CC -|- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
CC -|- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC PROTEIN FAMILY.

```
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb.sib.ch).
CC      -----
DR      EMBL; X81582; CAA57272.1; -
DR      EMBL; X76066; CAA53667.1; -
DR      EMBL; Z95492; CAB08859.1; -
DR      HSSP; P01180; INPO.
DR      MGD; MGI:96439; Igfbp4.
DR      InterPro; IPR000867; IGFBP.
DR      InterPro; IPR000716; Thyroglobulin_1.
DR      Pfam; PF00219; IGFBP; 1.
DR      Pfam; PF00086; thyroglobulin_1; 1.
DR      SMART; SM00121; IB; 1.
DR      SMART; SM00211; TY; 1.
DR      PROSITE; PS00222; IGF_BINDING; 1.
DR      PROSITE; PS00484; THYROGLOBULIN_1; 1.
DR      Growth factor binding; signal; Glycoprotein.
FT      SIGNAL      1      21      BY SIMILARITY.
FT      CHAIN      22      254      INSULIN-LIKE GROWTH FACTOR BINDING
FT      DOMAIN      196      245      PROTEIN 4.
FT      CARBOHYD      125      125      THYROGLOBULIN TYPE I.
FT      CONFLICT      4      5      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CONFLICT      13      13      FG -> CS (IN REF. 1).
FT      CONFLICT      56      56      A -> T (IN REF. 1).
FT      CONFLICT      67      67      C -> S (IN REF. 1).
FT      CONFLICT      67      67      C -> G (IN REF. 1).
SQ      SEQUENCE      254 AA; 27807 MW; 58EF89CB514AE17 CRC64;

Query Match      10.1%; Score 129.5; DB 1; Length 254;
Best Local Similarity 24.2%; Pred. No. 0.00012;
Matches 52; Conservative 34; Mismatches 76; Indels 53; Gaps 12;

OY      37 CQQLQASANSGL-----IGTVYPCKETGEFEKQOWSGTGYCWCVDDEGKEIL 85
DB      44 CEELVREPCCGCATCALGLGMPGCGVYTPRCG-----SGMRCYPPRGV-----EKPL 90
OY      86 GFKINGSPDCSRKALFLCQMMQALIVNPGWCGPSCKADGSPDEVQCCASNGECY-- 143
DB      91 RFLMHGQGYC-----TELSEIFAIOESLQTSKDESEHPNNSFN--PCSAHDHRCLOK 141
OY      144 ----CVDKKGELEGT-RQGRP----TCERHLSECEPARIKAHNSILRVEMF--VPEC 191
DB      142 HMAKIRDRKMKIYGTPREPRRPVPGSGQSELHRALE-RLAASQSRTHEDLFILIPNC 200
OY      192 LEDGSYNPYQWPS---TGycwCVD-EGYKVPg 221
DB      201 DRNGNHFPRQCHPALDGGRCWKCVDRKTYKLPg 235
```

Search completed: March 8, 2002, 09:25:29
Job time: 192 sec

4
1
1
1

1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2002, 09:24:02 ; Search time 23.53 Seconds

(Without alignments)
1435.993 Million cell updates/sec

Title: US-09-445-480A-2

Sequence: 1 MALSQNAQAFKSGFVMIWV.....VDEGVKVPQSDVRKRPPTC 231

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP_ARCHAEA: *
2: SP_BACTERIA: *
3: SP_FUNGI: *
4: SP_HUMAN: *
5: SP_INVERTEBRATE: *
6: SP_MAMMAL: *
7: SP_MHC: *
8: SP_ORGANELLE: *
9: SP_PHAGE: *
10: SP_PLANT: *
11: SP_RODENT: *
12: SP_VIRUS: *
13: SP_VERTEBRATE: *
14: SP_UNCLASSIFIED: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	1287	100.0	231	5	Q9U6K8
2	1278	99.3	231	5	Q9U4R8
3	1279	21.7	212	6	Q28330
4	259.5	20.2	446	21	Q9EP95
5	259	20.1	457	4	Q9H300
6	252.5	19.6	452	11	Q9H999
7	252.5	19.6	446	4	Q9H4P8
8	252.5	19.6	446	4	Q9H4P8
9	236.5	18.5	1161	5	Q9H901
10	177.5	13.8	688	5	Q98000
11	177.5	13.4	613	5	Q9V516
12	171.5	13.3	1511	5	Q9V821
13	167.5	13.0	126	4	Q9HCV2
14	167	13.0	234	7	Q9KRE4
15	167	13.0	237	13	Q9P0T2
16	166.5	12.9	279	4	Q9H4F7
17	165	12.8	2225	5	Q45891
18	161	12.5	439	4	Q08629
19	160.5	12.5	474	13	Q919K4

20	159.5	12.4	286	7	Q9BD54
21	159	12.4	442	11	Q62288
22	158.5	12.3	2150	5	Q44131
23	156	12.1	433	4	Q9B016
24	154	12.0	436	11	Q9ER59
25	153.5	11.9	198	13	Q9PVD9
26	153	11.9	436	4	Q75705
27	150	11.7	71	7	Q78208
28	145	11.3	424	4	Q92563
29	143	11.1	423	11	Q9ER58
30	143	11.1	660	5	Q9VCR3
31	128.5	10.0	505	5	Q25413
32	118.5	9.2	227	13	Q9PD12
33	114.5	8.9	1187	2	Q49549
34	114	8.9	911	11	Q9CRX6
35	112.5	8.7	226	5	Q44466
36	112	8.7	369	5	Q17457
37	111.5	8.7	949	5	P90956
38	110.5	8.6	291	6	Q9G1V5
39	110.5	8.6	434	5	Q9W0A0
40	109.5	8.5	293	6	Q9PT10
41	109.5	8.5	1069	5	Q9RPS2
42	108	8.4	955	4	Q9Y466
43	108	8.4	1284	5	Q9Y820
44	108	8.4	1664	5	Q9Y022
45	108	8.4	1999	4	Q99940

ALIGNMENTS

RESULT ID	1	PREDIMINARY:	PRF:	231 AA.
Q9U6K8	Q9U6K8			
AC	Q9U6K8			
DT	01-MAY-2000 (TREMBL)	13, Created		
DT	01-MAY-2000 (TREMBL)	13, Last sequence update		
DT	01-JUN-2001 (TREMBL)	17, Last annotation update		
DE	EQUISTATIN PRECURSOR.			
GN	EOL.			
OS	Actinia equina (European sea anemone).			
OC	Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;			
CC	Nynanthoeae; Actiniidae; Actinia.			
CX	NCBI_TaxID=6106;			
FN	11			
RP	SEQUENCE FROM N.A.			
RA	Galea K., Struckel B., Bayec S., Turk V., Lenarcic B.			
RT	"Molecular cloning and functional expression of equistatin."			
RL	Biol. Chem. 0:0-0(1999).			
DR	EMBL: AF184891; AAF02722.1; "			
DR	InterPro: IPR000716; Thyroglobulin_1.			
DR	Pfam: PF00086; thyroglobulin_1; 3.			
DR	PROSITE: PS00484; THYROGLOBULIN_1; 2.			
DR	SMART: SM00211; Ty; 3.			
KW	SIGNAL.			
FT	CHAIN	1	32	POTENTIAL.
FT	SIGNAL	33	231	EQUISTATIN.
SO	SEQUENCE	231 AA;	25406 MW;	B625D49272835958 CMC64;

Query Match 100.0%; Score 1287; DB 5; Length 231;
Best Local Similarity 100.0%; Pred. No. 6.1e-129; Indels 0; Gaps 0;
Matches 231; Conservative 0; Mismatches 0;

QY	1	MALSQNAQAFKSGFVMIWVLFACATITSTASITKCOLOLSANSGLITGYPOCKEKG 60
DB	1	MALSQNAQAFKSGFVMIWVLFACATITSTASITKCOLOLSANSGLITGYPOCKEKG 60
QY	61	EFEKCKWGSIGYCWCVDEKGRKILGTRKSPCSRRAALITLCOMMAIIVYWGK 120
DB	61	EFEKCKWGSIGYCWCVDEKGRKILGTRKSPCSRRAALITLCOMMAIIVYWGK 120
QY	121	PPSCRADSPFUEVQACASNGECYVDKRGKLEGTROGRPTCERHUSSECEARIRAHNS 180

```

DB 121 PSCADSGFDEVOCCASNGECYVCKKKELEGTRQOGRTCEHRLSECEBARIKASHN 180
QY 181 SLRYEMFVPECELEDSYNPVOCWPCSTGYCWCDEGCVKVPGSVDYRFRPTC 231
DB 181 SLRYEMFVPECELEDSYNPVOCWPCSTGYCWCDEGCVKVPGSVDYRFRPTC 231

RESULT 2
ID Q904R8 PRELIMINARY: PRT: 231 AA.
AC Q904R8;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DE 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE EQUITATIN PRECURSOR.
OS Actinia equina (European sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;
OC Nymphaeae; Actinidae; Actinia.
OX NCBI_Taxid=6106;
RN [1]
RP SEQUENCE FROM N.A.
RA Struckelj B., Pungercar J., Jongsma M.A.;
RT "Primary structure of equistatin deduced from cDNA.";
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF156179; AAF24173.1;
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF00086; thyroglobulin_1.
DR PROSITE: PS00484; THYROGLOBULIN_1; 2.
DR SMART: SM00211; TY; 3.
KW SIGNAL.
FT SIGNAL.
SQ SEQUENCE 231 AA; 25410 MW; 07E25687D3356E9 CRC64;

Query Match
Best Local Similarity 99.3%; Score 1278; DB 5; Length 231;
Matches 229; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MALSONAKFSKGFVIMVETFCATSTPEASTKCOLOASNSGLITGVPOCKETG 60
DB 1 MALSONAKFSKGFVIMVETFCATSTPEASTKCOLOASNSGLITGVPOCKETG 60
QY 61 EPEKQCGSTGTCMCVDEGKEILGTRKINGSPDCSRKKAALTICOMQAIIVNPGMG 120
DB 61 EPEKQCGSTGTCMCVDEGKEILGTRKINGSPDCSRKKAALTICOMQAIIVNPGMG 120
QY 121 PSCADSGFDEVOCCASNGECYVCKKKELEGTRQOGRTCEHRLSECEBARIKASHN 180
DB 121 PSCADSGFDEVOCCASNGECYVCKKKELEGTRQOGRTCEHRLSECEBARIKASHN 180
QY 181 SLRYEMFVPECELEDSYNPVOCWPCSTGYCWCDEGCVKVPGSVDYRFRPTC 231
DB 181 SLRYEMFVPECELEDSYNPVOCWPCSTGYCWCDEGCVKVPGSVDYRFRPTC 231

RESULT 3
ID Q28330 PRELIMINARY: PRT: 212 AA.
AC Q28330;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DE 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE DNA FOR THYROGLOBULIN (TG) (FRAGMENT).
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_Taxid=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=69207610; PubMed=2495821;
Van Ommen G.J.B., de Vrijlder J.J.M., Sterk A., Mercken L.O.Y.,

```

```

RA Arndberg A.C., Baas F.;
RT "Studies on the structures of the normal and abnormal goat
RT thyroglobulin genes.";
RL Biochimie 71:211-221(1989).
DR EMBL: X14997; CA433105.1;
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF00086; thyroglobulin_1; 2.
DR PROSITE: PS00484; THYROGLOBULIN_1; 2.
DR SMART: SM00211; TY; 2.
DR NON_TER 212
FT NON_TER 212
SQ SEQUENCE 212 AA; 23791 MW; 100325E08E9B9E8 CRC64;

Query Match
Best Local Similarity 21.7%; Score 279; DB 6; Length 212;
Matches 62; Conservative 18; Mismatches 60; Indels 18; Gaps 6;

QY 52 YVPOCKETGEFEKQCGSTGYCWCDEGKEILGTRKINGSPDCSRKKAALTICOMQAI 111
DB 48 YVPOCKEAGSFQYQCGCKDASGWCVDADGREYVSGRQGRP-----ACLSFCQLQKQ 102
QY 112 I-----VNVPGMGCPSPCKADSGFDEVOCCASNGECYVCKKKELEGTRQOGRT-CEH 165
DB 103 ILTSYINSTASTVLPQCQDSGDYSFVQCDLRRQCGCVADAGMAYCTGQGRPQPCR 162
QY 166 HLECEP-EARIKASHNLSRYEMFVPECELEDSYNPVOC 202
DB 163 ----SCIRNRRLHGVGDRSP---PQSPDCAFPRQOC 194

RESULT 4
ID Q69895 PRELIMINARY: PRT: 446 AA.
AC Q69895;
DT 01-MAR-2001 (TRENBLREL. 16, Created)
DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
DE 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE SECRETED MODULAR CALCIUM-BINDING PROTEIN 2 PRECURSOR.
GN SMOG2 OR SMOG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Vanname C., Gossling S., Maurer P.;
RT "Cloning of smog-2, a secreted modular calcium-binding protein.";
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
DE -1- SIMILARITY: TO EF-HAND FAMILY.
DR EMBL: AJ249901; CAC10355.1;
DR MGI: MGI:1929881; Smog2.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR002350; kazal.
DR Pfam: PF00036; ehfnd; 2.
DR Pfam: PF00050; kazal; 1.
DR Pfam: PF00086; thyroglobulin_1; 2.
DR SMART: SM00211; TY; 2.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR Calcium-binding; Signal.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1
SQ SEQUENCE 446 AA; 49818 MW; 6C3303AE7212CC2A CRC64;

Query Match
Best Local Similarity 20.2%; Score 259.5; DB 11; Length 446;
Matches 58; Conservative 30; Mismatches 86; Indels 27; Gaps 5;

QY 41 QASANGSLITGVYQCKETGEFEKQCGSTGYCWCDEGKEILGTRKINGSPDCSRK 99

```

```
Db 98 QEQARKFEQVFIPECNDGTYGSOVQCHSTYGCWCVTPNGRPISTGTAVAHKTPRCP--- 154
Qy 100 AALTLCQMMQAIIVNVPCMG-----PPSCKADGSFDEVOCCASNGEYCVDRKKGKE 151
Db 155 -----GSINKVYQREBAGAKADAAPALETPQGGDEEDIASRYPLIMTEQVYSRQ 205
Qy 152 LEGRQOGRTPCRHLSECEARIKAHNSILRVEFVPECELEDGSSYNPVOCWPTSTGYCWC 211
Db 206 NKTNNKNSASCDQEHQSALEBAKQPKNDN-----VVIPECAGHGLYKPVQCHPSTGYCWC 260
Qy 212 V-DEGVKVPVGSVDVRFKRPCT 231
Db 261 VLVDTPGRIPGTSTRYEQPKC 281

RESULT 5
Q9H3U0 PRELIMINARY; PRT; 457 AA.
AC Q9H3U0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SNAP-2B.
GN SNAP-2B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RA Nishimoto S., Toyoda H., Tawara J., Aoki T., Komurasaki T.;
RT "Molecular cloning and characterization of human smooth muscle cell
RT associated protein 2b(SNAP-2b).";
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO EF-HAND FAMILY.
DR EMBL: AB014737; BAB20274.1; -.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR002350; KAZAL.
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF00036; efhand; 2.
DR Pfam: PF00050; KAZAL; 1.
DR Pfam: PF00086; thyroglobulin_1; 2.
DR SMART: SM00280; KAZAL; 1.
DR SMART: SM00211; TY; 2.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
KW Calcium-binding.
SQ SEQUENCE 457 AA; 50862 MW; D20AE3BB82785925 CRC64;

Query Match 20.1%; Score 259; DB 4; Length 457;
Best Local Similarity 29.6%; Pred. No. 2.7e-19;
Matches 60; Conservative 32; Mismatches 91; Indels 20; Gaps 7;

Qy 41 QASANSGLIGTYVPCKETGEFEKQCGWSTGYCWCVDDEGKEITGKI-RGSPDC---- 95
Db 98 QEQARKFEQVFIPECNDGTYGSOVQCHSTYGCWCVTPNGRPISTGTAVAHKTPRCPGSGV 157
Qy 96 -----SRKKAALTLCQMMQAIIVNVPCMGPPSCKADGSFDEVOCCASNGEYCVDRKKGK 150
Db 158 NEKLPRREGTGKTV--SLQIFSVLNSDDAAPALETPQGGDEEDIASRYPLIMTEQVYSR 215
Qy 151 ELESTRQGRPTCER-ILHSCCEARIKAHNSILRVEFVPECELEDGSSYNPVOCWPTSTGYC 209
Db 216 Q-NKTNNKNSASCDQEHQSALEBAKQPKNDN-----VVIPECAGHGLYKPVQCHPSTGYC 269
Qy 210 WCV-DEGVKVPVGSVDVRFKRPCT 231
Db 270 WCVLVDTPGRIPGTSTRYEQPKC 292

RESULT 6
Q9WVN9
```

```
ID Q9WVN9 PRELIMINARY; PRT; 452 AA.
AC Q9WVN9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SPARC-RELATED PROTEIN.
GN SMOC1 OR SRG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SV;
RA Poliev A., Plachov D.;
RT "A Sparc-related gene (SRG).";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF070470; MAD41590.1; -.
DR HSSP: P09486; ISRA.
DR MGD: MGI:1929878; Smoc1.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR002350; KAZAL.
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF00050; KAZAL; 1.
DR Pfam: PF00086; thyroglobulin_1; 2.
DR SMART: SM00280; KAZAL; 1.
DR SMART: SM00211; TY; 2.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
DR PROSITE: PS00484; THYROGLOBULIN_1; 2.
DR PROSITE: PS00484; THYROGLOBULIN_1; 2.
SQ SEQUENCE 452 AA; 49763 MW; EAF73CB0AFC060F0 CRC64;
```

```
Query Match 19.9%; Score 255.5; DB 11; Length 452;
Best Local Similarity 32.0%; Pred. No. 6.3e-19;
Matches 71; Conservative 25; Mismatches 87; Indels 39; Gaps 8;

Qy 31 EASLTKCGO---QLQASANSGLIGTYVPCKETGEFEKQCGWSTGYCWCVDDEGKEITG 86
Db 88 DAGQSKRLERAOALEQAKKPOEAVVFPVCGEDGSTFOVCHTYGTCWCTPPDGRPISG 147
Qy 87 TKIRG-SPDCSRKKAALTLCQMMQAIIVNVPCMGPPSCKADGS-----FDEVO 134
Db 148 SSVQNTKTPVCSGVPTDKPLSQ-----GNSGRKDDGSKPTPTMETQPYFDDGE 194
Qy 135 CASNGEYCVDRKKGKELEGTROGGRP---TGER-ILHSCCEARIKAHNSILRVEFVPE 190
Db 195 ITAPFLIMIKHLVTKDSKLNNTVNRSEKVSQDOEROSALEBAR-----QNPREGIVPE 249
Qy 191 CLEDGSSYNPVOCWPTSTGYCWCY-DEGVKVPVGSVDVRFKRPCT 231
Db 250 CARGGLKFPVQCHQSTGYCWCYLVDTGRPLPGTSTRYVWVPC 291

RESULT 7
Q9H4F8 PRELIMINARY; PRT; 434 AA.
AC Q9H4F8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SECRETED MODULAR CALCIUM-BINDING PROTEIN PRECURSOR.
GN SMOC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Vannanthe C., Goesling S., Maurer P.;
RT "Cloning of human smoc-1, a novel secreted modular calcium-binding
RT protein.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
```


FT	REPEAT	276	306	SIMILARITY).
FT	REPEAT	550	605	EGF-LIKE.
FT	REPEAT	559	561	EGF-LIKE.
FT	REPEAT	562	564	1.
FT	REPEAT	565	567	2.
FT	REPEAT	568	570	3.
FT	REPEAT	571	573	4.
FT	REPEAT	606	664	5.
FT	REPEAT	665	704	EGF-LIKE.
FT	REPEAT	705	749	EGF-LIKE.
FT	REPEAT	1128	1161	EGF-LIKE.
FT	CARBOHYD	107	137	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	334	344	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	360	360	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	484	484	N-LINKED (GLCNAC. . .)
50	SEQUENCE	1161 AA:	128471 MW:	(POTENTIAL).

	Query Match	18.5%;	Score 238.5;	DB 5;	Length 1161;	
	Best Local Similarity	32.3%;	Pred. NO. 1.1e-16;			
	Matches	72;	Conservative	31;	Mismatches	85;
					Indels	37;
					Gaps	13.
Oy	28	TSTESLTKRCQ-QLOASANSGLIG-----TYVPOCKETGEFEERKOCMGSTGYCMCYDED	80			
Dd	595	TFTE-PMYVCEREHLEAEEMRDLGKEINDELFFPCTNLGTFRAPROQTTHYDSVCMDPD	653			
Oy	81	-GKELTGKRINSIP-----DCSRKK---AALTLCOMAOALIVNPGMCGP-----PS	123			
Dd	654	SGVEIDGT--KSLPREGYLDCSDVRPPSPGOTPCCEERTIENMRNRAGSEANSIEMFR	711			
Oy	124	CKADSPEDVOCCASNGEYCVDKKGKKELEGRRGOGRPLCERHLSECEARLKANSLS-	182			
Dd	712	CMESSATPEPVQ-QLPDGTIFMCYDFEGEELSREEE--PRC---LTSCOLAIYALESNNITF	765			
Oy	183	RVEMFVEPECTLEDGSYNPYOVCSPTGYCMCYDEGGVKYPGSDV	224			
Dd	766	KDTSLFYVQCCKPDGEYLPQQVAS--YGCVDNRNGLEISNSRV	806			
	RESULT	10				
ID	O98UD0	PRELIMINARY;	PRT;	88 AA.		
AC	O98UD0:					
Df	01-JUN-2001 (TREMBLrel. 17, Created)					
Df	01-JUN-2001 (TREMBLrel. 17, Last sequence update)					
Df	01-JUN-2001 (TREMBLrel. 17, Last annotation update)					
Dt	OVARIAN CYSTEINE PROTEASE INHIBITOR.					
OS	Capri.					
OS	Salvelinus fontinalis (Brook trout) (Brook char).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei;					
OC	Protacanthopterygii; Salmoniformes; Salmonidae; Salvelinus.					
OX	NCBI_TaxID=8038;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	Bobe J., Goetz F.W.;					
RT	"Identification of a cysteine protease inhibitor mRNA in the ovulatory					
RL	and postovulatory brook trout ovary."					
DR	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.					
RM	EMBL; AF223387; AAK00216.1; -					
KN	Protease.					
SQ	SEQUENCE	88 AA;	9171 MW;	F14E73AC2884F9E8 CRC64;		
	Query Match	13.8%;	Score 177.5;	DB 13;	Length 88;	
	Best Local Similarity	38.2%;	Pred. No. 2.1e-11;			
	Matches	29;	Conservative	18;	Mismatches	26;
					Indels	3;
					Gaps	1.
Oy	16	VMIWVFICAITTSFASL---TKCOQLDASANSGLIGTVVPQCKRTGTFEERKOCMGSGSNG	72			
Dd	3	LITLILVSATAALGATTPRCRRADAATHGPISGAVIPFCDAAGYTTPRKOCMGSSAG	62			

Oy	73	YCMCVDGCKEILGK 88		1:1	11:
Db	63	YCMCVSTGCKIOGTE 78			
RESULT	11				
09V5F6	ID	PRELIMINARY;	PRT:	613	AA.
AC	09V5F6				
DT	01-MAY-2000 (TREMblrel_13, Created)				
DT	01-May-2000 (TREMblrel_13, Last sequence update)				
DT	01-JUN-2001 (TREMblrel_17, Last annotation update)				
CC	CG2264	PROTEIN.			
DE	CG2264				
GN	CG2264				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OX	Ephydrioidae; Drosophilidae; Drosophila.				
NC	NCBI_TaxID=7227;				
NP	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPRAIN-BREKLEY;				
RA	MEDLINE-20196006; PubMed-10731132;				
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Bridton R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,				
RA	Wan K.H., Doyle C., Baxter E.G., Holt G., Nelson C.R., Miklos G.L.G.,				
RA	Abtill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,				
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beckson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,				
RA	Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,				
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dubin K.J., Doup L.E., Downes W., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Dudson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,				
RA	Foster C.A., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,				
RA	Gloder K., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,				
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,				
RA	Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,				
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,				
RA	Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,				
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,				
RA	Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,				
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui Y., Reese M.G.,				
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,				
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,				
RA	Spier E., Spradling A.C., Stappleton M., Strong R., Sun E.,				
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,				
RA	Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weisenbach J.,				
RA	Williams S.W., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.-A.,				
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,				
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,				
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,				
RT	"The genome sequence of <i>Drosophila melanogaster</i> ."				
RL	Science 287:2185-2195(2000).				
CC	-1- SEMIARITY: TO EE-HAND FAMILY.				
DR	EMBL: AE003831; AAF58854.1; -				
DR	HSSP: P80424; IANI				
DR	FLYBase: FBgn0033478; CG2264.				
DR	InterPro: IPR002048; EF-hand.				
DR	InterPro: IPR002350; Kazal.				
DR	InterPro: IPR00716; Thyroglobulin_1.				
DR	Pfam: PF00036; ehand; 2.				
DR	Pfam: PF00086; thyroglobulin_1; 2.				
DR	SMART: SM00280; KAZAL; 1.				
DR	SMART: SM00211; TY; 2.				

```

DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
DR PROSITE: PS00484; THYROGLOBULIN_1; 1.
KW Calcium-binding.
SQ SEQUENCE 613 AA; 68955 MW; DE7EDC32B73E8A1A CRC64;

Query Match
Best Local Similarity 13.4%; Score 173; DB 5; Length 613;
Matches 70; Conservative 37; Mismatches 74; Indels 188; Gaps 14;

OY 50 GTVPCKETGTEFEERKCGSTGYCVCVEDEKEI----- 84
| : : : : | : : : : | : : : : | : : : : |
DB 96 GTVPCKCRDGNFAMQCTGNNG-CWCSDSQGRPIADNDKQPRRRKGLCRANRRDRRL 154
| : : : : | : : : : | : : : : | : : : : |
OY 85 LGTKIRGSPD-----CSR-----RKAALT-- 103
| : : : : | : : : : | : : : : | : : : : |
DB 155 ASHQIGYNPDTSASKGSSSAGSTARHTCKSRDSQENTNLAMFENASFFRQSLSDS 214
| : : : : | : : : : | : : : : | : : : : |
OY 104 -----LQOMQAIIVN-PGWGPSPCK-----ADGSPDEV 133
| : : : : | : : : : | : : : : | : : : : |
DB 215 HLEWQFSLDTNGNKLDRQEIRELKVLRRNVRPRGRTFGKYCDVTKANLWLEW 274
| : : : : | : : : : | : : : : | : : : : |
OY 134 OCCAS-----NGECYC 144
| : : : : | : : : : | : : : : | : : : : |
DB 275 SVCFTRKRFHNRSAVYNLLASSAATAPRPHSTHYSHNTNVNGHHRGHTNTIGTVNPHS 334
| : : : : | : : : : | : : : : | : : : : |
OY 145 VDKKREL-----EGTRQGR--PTCERHL-----S 168
| : : : : | : : : : | : : : : | : : : : |
DB 335 EDASSEHEHDNYEDSATCYGEEDDSQCADLPSSRTIIPSLIYMLNSKPETASODLEND 394
| : : : : | : : : : | : : : : | : : : : |
OY 169 EC---EARIKASHNSLRVEMVPECLDEGSYNPQCM--PSTGYCWCVDG-GVKVPGS 222
| : : : : | : : : : | : : : : | : : : : |
DB 395 NCWMDQSVYLEQGHGKGSVLEFPQCLPDGRYQRIQCYGYSSTSTSCWCVNEDTGSINGT 454
| : : : : | : : : : | : : : : | : : : : |
OY 222 DVFRKRPC 231
| : : : : | : : : : | : : : : | : : : : |
DB 455 SVKNRPPC 463
| : : : : | : : : : | : : : : | : : : : |

RESULT 12
O9VB21 PRELIMINARY; PRT; 1511 AA.
AC O9VB21.
DT 01-MAY-2000 (TREMELREL. 13, Created)
DT 01-MAY-2000 (TREMELREL. 13, Last sequence update)
DT 01-JUN-2001 (TREMELREL. 17, Last annotation update)
DE CG5639 PROTEIN.
GN CG5639
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
| : : : : | : : : : | : : : : | : : : : |
RN SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chertiy J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelle S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

```

```

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodali C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "the genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003761; AAF56722.1; -.
DR HSSP; P15358; 1SKZ.
DR FLYBASE; FBgn0039527; CG5639.
DR InterPro; IPR002899; EB.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR003645; FOIN.
DR InterPro; IPR002223; Kunitz_BPT1.
DR InterPro; IPR000716; Thyroglobulin_1.
DR InterPro; IPR002221; WAP.
DR InterPro; IPR001680; WDA0.
DR Pfam; PF00086; thyroglobulin_1; 5.
DR Pfam; PF00095; wap; 2.
DR PRINTS; PR00003; 4DISULPHORE.
DR SMART; SM00274; FOIN; 3.
DR SMART; SM00131; KU; 1.
DR SMART; SM00211; TY; 5.
DR SMART; SM00217; WAP; 2.
DR SMART; SM00289; WRI; 2.
DR PROSITE; PS00317; 4_DISULFIDE_CORE; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_3.
DR PROSITE; PS00484; THYROGLOBULIN_1; 3.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
SQ SEQUENCE 1511 AA; 165894 MW; BD44D64CB9DCC3B8 CRC64;

```

```

Query Match
Best Local Similarity 13.3%; Score 171.5; DB 5; Length 1511;
Matches 72; Conservative 36; Mismatches 95; Indels 163; Gaps 17;

OY 25 CAITSEIASI-TRQDQLA-----SANSGLIGTY--VPQC-KETGEFEERKCGSTGYCW 75
| : : : : | : : : : | : : : : | : : : : |
DB 823 CGTQCVDPOLKTRACQHIQAIOLHSSSELTIPAROMAVACDPNNGKMNQVQC-SPDGHCW 881
| : : : : | : : : : | : : : : | : : : : |
OY 76 CVDEDEKELIGTKIR-----GSPGCSRRKAL 102
| : : : : | : : : : | : : : : | : : : : |
DB 882 CVDQDCKILPGRVAKSPATPKCOENSSFPACPTNCSLECGSYQMDNSNCPCECRNYCN 941
| : : : : | : : : : | : : : : | : : : : |
OY 103 TL-----COMQAIIVNP-----GWCSP 121
| : : : : | : : : : | : : : : | : : : : |
DB 942 EVASCPHEHCQILSVSCVSPCKMPICVPRRASITCPGCPNPLQGDGLDMSCGPHNEHYC 1001
| : : : : | : : : : | : : : : | : : : : |
OY 122 ---PSCKADGSPDEVQCCASNGE-CY-----CVDKKELEGTR----- 156
| : : : : | : : : : | : : : : | : : : : |
DB 1002 PTHSCOLPVPNNRVCCKSKTRDVCFESMDNCGLATGKSEKSTRTSPRSPKANKCLPVI 1061
| : : : : | : : : : | : : : : | : : : : |
OY 157 QGGRPTCERH-----LSDEEARIKASHNSLRVE-----MFVEEC-LE 193
| : : : : | : : : : | : : : : | : : : : |
DB 1062 DSAAPACQTKNLPHNELACNSCPVLTQCERLIK-NNLAAORTGSSVWFQRCBPV 1118
| : : : : | : : : : | : : : : | : : : : |
OY 194 DGSYNPQCM-----PSNGYCMCVDEGSGVKVPGSPVR 225
| : : : : | : : : : | : : : : | : : : : |

```


Db 1119 TGHMSPVQCLGKOPQPMDRHTEIVSRAFASEAPASAGEAPGVWCADKKGAPLKGTLTR 1178
 OY 226 FKRPCT 231
 Db 1179 ESEPIC 1184

RESULT 13

O9HCY2 PRELIMINARY; PRT; 126 AA.

AC O9HCY2;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE D421D16.1 (NOVEL THYROGLOBULIN TYPE-1 REPEAT CONTAINING PROTEIN)
 DE (FRAGMENT)
 GN D421D16.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Sycamore N.
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL109940; CAC03594.1; -
 DR InterPro: IPR002350; kazal.
 DR InterPro: IPR000716; Thyroglobulin_1.
 DR Pfam: PF00050; kazal; 1.
 DR Pfam: PF00086; Thyroglobulin_1; 1.
 DR SMART: SM00280; KAZAL; 1.
 DR SMART: SM00211; TY; 1.
 DR PROSITE: PS00484; THYROGLOBULIN_1; 1.
 FT NON_TER 1 1
 FT NON_TER 126 126
 SQ SEQUENCE 126 AA; 14281 MW; D7076BEF2B89ABDA CRC64;

Query Match 13.0%; Score 167.5; DB 4; Length 126;
 Best Local Similarity 25.6%; Pred. No. 3.6e-10;

Matches 43; Conservative 17; Mismatches 49; Indels 59; Gaps 4;

OY 77 VDEGKEILGTIRGSPDCSRRAALTLQMMQAIIVNVEGCGPSCADG-----S 129
 Db 4 VDDKDKDCLDCAGSPQ-----KPLCAGDGRTPLSRCE 37
 OY 130 FDEYQC-----CASNECCVDKKGKKELEGTROGGRPTGERHLSCEEFARIAHSNSLR 183
 Db 38 FORAKCKDPLEIAIYRNCNCDVSRCAERYTQEOARKERO-----S 78
 OY 184 VEMFVPCLEBDGSYNPQCPSPSTGYCWCVDGEGVAVKPGSDVRFRPRTC 231
 Db 79 -QVPIPCNDGTYSOVQCHSYTCYCWCTPNGRPISTGIVAAKTPRC 125

RESULT 14

O9XREA PRELIMINARY; PRT; 234 AA.

AC O9XREA;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE MHC CLASS II-ASSOCIATED INVARIANT CHAIN.
 OS Brachydanio rerio (zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Rasbora; Danio.
 OX NCBI_Taxid=7955;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-EX; TISSUE=SPLEEN;
 RA Natarajan K., Castellino F., Germain R.N., Margulies D.H.;
 RT "The zebrafish invariant chain promotes the transport and maturation

RT of mouse MHC-II molecules.*;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF114830; AAD24542.1; -

DR InterPro: IPR000716; Thyroglobulin_1.
 DR Pfam: PF00086; Thyroglobulin_1; 1.
 DR PROSITE: PS00484; THYROGLOBULIN_1; UNKNOWN_1.
 DR SMART: SM00211; TY; 1.
 SQ SEQUENCE 234 AA; 25916 MW; 76C5917D7DFA6362 CRC64;

Query Match 13.0%; Score 167; DB 7; Length 234;
 Best Local Similarity 50.8%; Pred. No. 8e-10;

Matches 31; Conservative 5; Mismatches 23; Indels 2; Gaps 1;

OY 35 TKGQQLQASANGSLGTYVPOCKENGFEFEKQCMSTGCMCVDDGKEILGTIRGSPD 94
 Db 176 TKCK--LSESEVRRGFEFPACDEBGNVPMQCMHSTGCMCVTRDGEIESTIRGRPQ 233
 OY 95 C 95
 Db 234 C 234

RESULT 15

O9PUT2 PRELIMINARY; PRT; 237 AA.

AC O9PUT2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE INVARIANT CHAIN-LIKE PROTEIN 1.
 GN ICLP-1.
 OS Brachydanio rerio (zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Rasbora; Danio.
 OX NCBI_Taxid=7955;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE-20009532; PubMed=10541812;
 RA Yoder J.A., Haire R.N., Litman G.W.;
 RT "Cloning of two zebrafish cDNAs that share domains with the MHC class II associated invariant chain."
 RL Immunogenetics 50:84-88(1999).
 DR EMBL; AF148214; AAF01568.1; -
 DR InterPro: IPR000716; Thyroglobulin_1.
 DR Pfam: PF00086; Thyroglobulin_1; 1.
 DR PROSITE: PS00484; THYROGLOBULIN_1; UNKNOWN_1.
 DR SMART: SM00211; TY; 1.
 SQ SEQUENCE 237 AA; 26231 MW; E56D326C5917D7D CRC64;

Query Match 13.0%; Score 167; DB 13; Length 237;
 Best Local Similarity 50.8%; Pred. No. 8.1e-10;

Matches 31; Conservative 5; Mismatches 23; Indels 2; Gaps 1;

OY 35 TKGQQLQASANGSLGTYVPOCKENGFEFEKQCMSTGCMCVDDGKEILGTIRGSPD 94
 Db 176 TKCK--LSESEVRRGFEFPACDEBGNVPMQCMHSTGCMCVTRDGEIESTIRGRPQ 233
 OY 95 C 95
 Db 234 C 234

Search completed: March 8, 2002, 09:25:59
 Job time: 117 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2002, 09:21:52 ; Search time 17.1 Seconds
(without alignments)
1029.025 Million cell updates/sec

Title: US-09-445-480a-2

Perfect score: 1287

Sequence: 1 MALSONQAKFKSGFYVMIV.....VDEGVKVPESDVREKRPIC 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	295	22.9	2769	1 UIBO	thyroglobulin prec
2	279.5	21.7	243	2 A36124	thyroglobulin 2 pr
3	279	21.2	212	2 I46176	thyroglobulin prec
4	271	21.1	2767	1 UIHU	thyroglobulin prec
5	245.5	19.1	1376	2 G00043	osteonidogen - hum
6	238.5	18.5	1161	1 S31213	nidogen precursor
7	165	12.8	2225	2 T26063	hypothetical prote
8	159	12.4	439	2 S33293	testican - human
9	158.5	12.3	2150	2 T32497	hypothetical prote
10	154	12.0	1245	1 MMMSND	nidogen precursor
11	153	11.9	280	2 S04362	class II histocomp
12	153	11.9	1247	1 MMHUND	nidogen precursor
13	150	11.7	71	2 I70687	protein 41 - human
14	135	10.5	279	1 HLHMSG	class II histocomp
15	132.5	10.3	258	2 A45403	insulin-like growt
16	129.5	10.1	254	2 I48599	insulin-like growt
17	129	10.0	254	2 I48603	insulin-like growt
18	128.5	10.0	254	2 JC1464	insulin-like growt
19	124.5	9.7	237	2 I47031	insulin-like growt
20	122.5	9.5	258	1 B37552	insulin-like growt
21	122.5	9.5	271	2 JC4584	insulin-like growt
22	122.5	9.5	272	2 A53748	insulin-like growt
23	120.5	9.4	271	2 I48604	insulin-like growt
24	118.5	9.2	157	2 A25604	thyroglobulin - ra
25	118.5	9.2	271	2 JC1465	insulin-like growt
26	114.5	8.9	1167	2 T18355	hypothetical prote
27	114	8.9	263	1 S23009	insulin-like growt
28	114	8.9	1786	1 MMMSB1	lamin beta-1 cha
29	113.5	8.8	240	2 A39842	insulin-like growt

30	112	8.7	369	2 T29836	hypothetical prote
31	112	8.7	1786	1 MMHNB1	lamin beta-1 cha
32	111.5	8.7	949	2 T24294	hypothetical prote
33	110	8.5	266	1 A35037	insulin-like growt
34	110	8.5	291	1 JN0064	insulin-like growt
35	109.5	8.5	259	1 IOH01	insulin-like growt
36	109.5	8.5	4544	1 S02392	alpha-2-macroglobu
37	109	8.5	291	2 I48602	insulin-like growt
38	108.5	8.4	252	1 A36748	insulin-like growt
39	108	8.4	212	1 A36082	insulin-like growt
40	108	8.4	1620	2 T27283	hypothetical prote
41	107.5	8.4	623	2 T47542	Spot 3 protein and
42	107	8.3	226	2 JN0464	insulin-like growt
43	107	8.3	238	2 I48605	insulin-like growt
44	107	8.3	291	1 IOH03	insulin-like growt
45	107	8.3	305	2 I48601	insulin-like growt

ALIGNMENTS

RESULT 1

UIBO

C:Species: Bos primigenius taurus (cattle)
C:Date: 28-May-1986 #sequence.revision 28-May-1986 #text.change 16-Jun-2000
C:Accession: A01533; S10078; S05566; A32735; S39431; A32740; S29734; I45981
R:Mercken, L.; Simons, M.J.; Swillens, S.; Masset, M.; Vassart, G.
N:ature 316, 647-651, 1985
A:Title: Primary structure of bovine thyroglobulin deduced from the sequence of 1
A:Accession: A01533
A:Reference number: A01533; MUID:85296288

A:Molecule type: mRNA

A:Residues: 1-2769 <MER>

A:Cross-references: GB:X02815; GB:J00040; NID:9774; PIDN:CA26584.1; PID:9775
R:Hansen, C.; Gerard, C.; Vassart, G.; Stordeur, P.; Christophe, D.
Eur. J. Biochem. 178, 387-393, 1988

A:Title: Thyroid-specific and CAMP-dependent hypersensitive regions in thyroglobu
A:Reference number: S10078; MUID:89091173

A:Accession: S10078

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-20 <HAN>

A:Cross-references: EMBL:X06071; NID:9788; PIDN:CA32504.1; PID:9789
R:Parma, J.; Christophe, D.; Pohl, V.; Vassart, G.
J. Mol. Biol. 196, 769-779, 1987

A:Title: Structural organization of the 5' region of the thyroglobulin gene. Evic
A:Reference number: S03422; MUID:88062712

A:Accession: S03566

A:Molecule type: DNA

A:Residues: 1002-1205, 'R', 1207-1209 <PAR>

A:Cross-references: EMBL:X06071; NID:9778; PIDN:CA29457.1; PID:92204111
R:Rawlitch, A.B.; Liltner, M.R.; Gregg, J.; Turner, C.D.; Rouse, J.B.; Hamilton, J.
Biochem. Biophys. Res. Commun. 118, 423-429, 1984

A:Title: The isolation of identical thyroxine containing amino acid sequences fro
A:Reference number: A32735; MUID:84153804

A:Accession: A32735

A:Molecule type: Protein

A:Residues: 20-38 <RAW>

R:gentile, F.; Salvatore, G.
Eur. J. Biochem. 218, 603-621, 1993

A:Title: Preferential sites of proteolytic cleavage of bovine, human and rat thyr
A:Reference number: S39431; MUID:94094855

A:Accession: S39431

A:Molecule type: Protein

A:Residues: 20-27;259-265;521-526, 'X', 528;539-560;1161-1168, 'XX', 1171-1172;1203-11
R:Rawlitch, A.B.; Mercken, L.; Hamilton, J.W.; Vassart, G.
Biochem. Biophys. Res. Commun. 119, 335-342, 1984

A:Title: The structure of a naturally occurring 10K polypeptide derived from the
A:Reference number: A32740; MUID:84153919

A:Accession: A32740

A:Molecule type: Protein

A:Residues: 20-38;66-78;95-99 <RA2>

R:Rawlitch, A.B.; Pollock, H.G.; Yang, S.X.
Arch. Biochem. Biophys. 300, 271-279, 1993
A:Title: Thyroglobulin glycosylation: location and nature of the N-linked oligosaccharide
A:Reference number: 529734; MUID:93145326
A:Accession: 529734
A:Molecule type: protein
A:Residues: 106-109, 'X', 111-117:475-477, 'N', 479-482, 'X', 484-485;492-494, 'X', 496-502;839-252-2255;2272-2281 <R3>
R:Rickerts, M.H.; Simons, M.
Proc. Natl. Acad. Sci. U.S.A. 84, 3181-3184, 1987
A:Title: A nonsense mutation causes hereditary goitre in the Afrikaner cattle and unmass
A:Reference number: 145981; MUID:87204101
A:Accession: 145981
A:Status: Preliminary: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 712-723 <R1C>
A:Cross-references: GB:M16448; NID:g163743; PIDN:AAA30777.1; PID:g552340
C:Comment: The thyroglobulin molecule is a dimeric glycoprotein specific to the thyroid
C:Genetics:
A:Introns: 1047/1; 1073/1; 110/3; 1145/1
C:Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat homol
C:Keywords: duplication; glycoprotein; homodimer; Iodine; thyrold gland; thyrold hormone
E:1-19/Domaln: signal sequence #status predicted <SIG>
E:20-2769/Product: thyroglobulin #status predicted <TG>
E:34-92/Domaln: thyroglobulin type I repeat homology <TH01>
E:96-160/Domaln: thyroglobulin type I repeat homology <TH02>
E:166-297/Domaln: thyroglobulin type I repeat homology <TH03>
E:301-358/Domaln: thyroglobulin type I repeat homology <TH04>
E:607-657/Domaln: thyroglobulin type I repeat homology <TH05>
E:661-725/Domaln: thyroglobulin type I repeat homology <TH06>
E:729-921/Domaln: thyroglobulin type I repeat homology #status atypical <TH07>
E:925-1073/Domaln: thyroglobulin type I repeat homology #status atypical <TH08>
E:1077-1145/Domaln: thyroglobulin type I repeat homology <TH09>
E:1149-1210/Domaln: thyroglobulin type I repeat homology <TH10>
E:1458-1471/Region: type II repeat
E:1472-1488/Region: type II repeat
E:1489-1505/Region: type II repeat
E:1605-1725/Region: type III repeat
E:1726-1893/Region: type III repeat
E:1894-1996/Region: type III repeat
E:1997-2130/Region: type III repeat
E:2131-2188/Region: type III repeat
E:2229-2727/Domaln: cholinesterase homology <ACE>
E:24/Modified site: thyroxine (Tyr) #status experimental
E:110/Binding site: carbohydrate (Asn) (covalent) #status absent
E:199/Binding site: carbohydrate (Asn) (covalent) #status absent
E:2574/Modified site: thyroxine (Tyr) #status predicted
E:258/Modified site: thyroxine (Tyr) #status predicted
E:2767/Modified site: triiodothyronine (Tyr) #status predicted

Query Match	22.9%	Score 295	DB 1	length 2769
Best Local Similarity	33.5%	Pred. No. 3,1e-16		
Matches	69	Conservative	21	Mismatches 64; Indels 52; Gaps 8
QY	52	YVPCKETGEFEERKQCGWSTGYCWCVDDEKDEKITGRTKIRGSPDCSRKKAALFLCQMMQAI	111	
Db	48	YVPCADGEDGSFQVOCGCKDGASQWCVCADALGAREVPESRQGP-----AACLSFQQLQKQ	102	
QY	112	I-----VNVGWCGPCSKADGSFDEVDVOCASNGEBCYCVDRKGELEGTROGQGPPT-CER	165	
Db	103	ILUSSYINSTATSTYLPQCCDSGDYSVQCDLRRKQCCVDAGMVGITRQGRPARCPR	162	
QY	166	HLSECF-EARIKASHLSLRYEMFVEPCELEDGSGSYNVQC-----	202	
Db	163	---SCEIRNRRLHLHGVDGRSP---PQCSPDGAFRVPQCKLVNTDDMMIFDLVHSYRFPD	216	
QY	203	-----WPS-TGYCWCVDDEG	216	
Db	217	AEVTFSFSFRSFRPEVSGYCCYCCASQ	242	

```

RESULT      2
A36124
thyroglobulin 2 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 14-Dec-1990 #sequence_revision 14-Dec-1990 #text_change 11-Apr-1997
C:Accession: A36124
R:Graves, P.N.; Daviles, T.F.
MOL: Endocrinol. 4, 135-161, 1990
A:Title: A second thyroglobulin messenger RNA species (rTg-2) in rat thyrocytes.
A:Reference number: A36124; MUID: 90220642
A:Accession: A36124
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-243 <GRA>
A:Cross-references: GB:M35965
C:Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat
C:Keywords: dimer; glycoprotein; iodine; thyroid gland; thyroid hormone biosynthesis
F:35-93/Domains: thyroglobulin type I repeat homology <THY1>
F:97-161/Domains: thyroglobulin type I repeat homology <THY2>
F:25/Modified site thyroxine (Tyr) #status predicted

```

[illegible]

```

RESULT      3
146176
C: thyroglobulin precursor - goat (fragment)
C: Species: Capra aegagrus hircus (domestic goat)
C: Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 16-Jul-1999
C: Accession: I46176
R: Van Ommen, G.J.; Sterk, A.; Mercken, L.O.; Arnborg, A.C.; Baas, F.; De Vrijder,
Biochimie 71, 211-221, 1999
A: Title: Studies on the structures of the normal and abnormal goat thyroglobulin g
A: Reference number: I46176; MUID:89207610
A: Accession: I46176
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-212 <N>
A: Cross-references: EMBL:X14997; NID:9989; PTDN:CAA33105.1; PID:9990
C: Genomics:
A: Gene: TG
C: Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat
F: 34-92/Domain: thyroglobulin type I repeat homology <TR1>
F: 96-160/Domain: thyroglobulin type I repeat homology <TR2>

```

Query Match	21.7%	Score 279;	DB 2;	Length 212;
Best Local Similarity	39.28%	Pred. No. 6,6e-16;		
Matches	62;	Conservative	18;	Mismatches 60; Indels 18; Gaps 6;
0y	52	YVPCKCTGEEFEKQCGSGTGCWCVEVDGKELTATKRGSPDSCRRAALTLGQMOMAI	111	
db	48	YVPQCADGSGFQVVOGCKDGACSCWCVDADGRVPSRPQNP----	ACLSTCLQKQ00	102

DNA 8, 581-594, 1989
 A:Title: Human nidogen: complete amino acid sequence and structural domains deduced from
 A:Reference number: A33322; M0ID:90091745
 A:Accession: A33322
 A:Molecule type: mRNA
 A:Residues: 1-1247 <NAG>
 A:Cross-references: EMBL:M30269
 R:Olson, D.R.; Nagayoshi, T.; Farlo, M.; Mattel, M.G.; Passage, E.; Well, D.; Timpl, R.;
 Am. J. Hum. Genet. 44, 876-885, 1989
 A:Title: Human nidogen: cDNA cloning, cellular expression, and mapping of the gene to ch
 A:Reference number: A32437; M0ID:89270475
 A:Accession: A32437
 A:Molecule type: mRNA
 A:Residues: 667-1247 <OLS>
 A:Cross-references: EMBL:M27445; NID:g602466; PIDN:AAA57261.1; PID:g602467
 A:Note: the authors translated the codon AAG for residue 966 as Cys
 R:Farlo, M.J.; O'leary, J.; Kaehehl, V.M.; Chen, Y.Q.; Saitta, B.; Ulteo, J.
 J. Invest. Dermatol. 97, 281-285, 1991
 A:Title: Human nidogen gene: structural and functional characterization of the 5'-flank
 A:Reference number: A61367; M0ID:91302882
 A:Accession: A61367
 A:Molecule type: DNA
 A:Residues: 1-28 <FAZ>
 C:Comment: This protein is a basement membrane glycoprotein that forms a complex with la
 C:Genetics:
 A:Gene: GDB:NID
 A:Cross-references: GDB:120236; OMIM:131390
 A:Map position: 1q43-1q43
 C:Superfamily: 1q43-1q43
 C:Keywords: basement membrane; beta-hydroxyasparagine; calcium binding; cell binding; cd
 protein
 F:1-28/Domain: signal sequence #status predicted <STC>
 F:29-1247/Product: nidogen #status predicted <MAT>
 F:390-425/Domain: EGF homology <EG1>
 F:672-708/Domain: EGF homology <EG2>
 F:702-704/Region: cell attachment (R-G-D) motif
 F:714-750/Domain: EGF homology <EG3>
 F:762-800/Domain: EGF homology <EG4>
 F:806-839/Domain: EGF homology <EG5>
 F:849-919/Domain: thyroglobulin type I repeat homology <THY1>
 F:990-1032/Domain: LDL receptor WTD-containing repeat homology <YW1>
 F:1033-1075/Domain: LDL receptor WTD-containing repeat homology <YW2>
 F:1076-1120/Domain: LDL receptor WTD-containing repeat homology <YW3>
 F:1121-1160/Domain: LDL receptor WTD-containing repeat homology <YW4>
 F:1161-1197/Domain: LDL receptor WTD-containing repeat homology <YW5>
 F:1212-1243/Domain: EGF homology <EG6>
 F:289-296/Binding site: sulfate (Tyr) (covalent) #status predicted
 F:729, 819/Modified site: eythro-beta-hydroxyasparagine (asn) #status predicted
 F:756/Cross-link: isopeptide (Gln) (interchain to Lys N6-amino of laminin) #status predi
 F:1137/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 11.9%; Score 153; DB 1; Length 1247;
 Best Local Similarity 25.5%; Pred. No. 6.8e-05;
 Matches 56; Conservative 18; Mismatches 70; Indels 76; Gaps 11;
 Oy 25 CAITSTASTLKCOQLASNSGLIGTYVP--QCKETGEFEKCKGSGTCYCMCV--D 78
 Db 750 CVAAVVDORPIWYCE-----TGJHNCDDIPQRAQCIYTG-----GSSYTCCLPGRS 794
 Oy 79 EDGKEILGTRKRGSPDCSRKKAALTLCCMMQALIVNPGCGSPCKADGSPFEVCCAS 138
 Db 795 GDGQ-----ACQDQDE-----CQPSRCHPAF-----CYWT 820
 Oy 139 NGECYCYDKKKELEGTROOGRPTC--ERNHLSCEERAKIKH-----SNSLRVE 185
 Db 821 PGSTFCQCKRPGYDGR-----CVPGEVKRRCQHER--EHLILAAGATDQRRIIPGG 872
 Oy 186 MFVPECLEDEGSYNPVQCPSTGYCMCYDEGCVVPPGSDVR 225
 Db 873 LFVPECDAGHYAPTOCHGSGTCYCMCVDRDRGREVEGRTTR 912

RESULT 13
 I70687
 Protein 41 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 23-Jul-1999
 C:Accession: I70687
 R:O'Sullivan, D.M.; Noonan, D.; Quaranta, V.
 J. Exp. Med. 166, 444-460, 1987
 A:Title: Four Ia invariant chain forms derive from a single gene by alternate spl
 A:Reference number: I55653; M0ID:87252940
 A:Accession: I70687
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-71 <RES>
 A:Cross-references: GB:M28590; NID:g292154; PIDN:AAA35996.1; PID:g292155
 C:Superfamily: class II histocompatibility antigen-associated gamma chain; thyrog
 F:1-66/Domain: thyroglobulin type I repeat homology <THY1>

Query Match 11.7%; Score 150; DB 2; Length 71;
 Best Local Similarity 42.9%; Pred. No. 9.5e-06;
 Matches 27; Conservative 12; Mismatches 24; Indels 0; Gaps 0;
 Oy 34 LTKCOQLASNSGLIGTYVPCKETGEFEKCKGSGTCYCMCVDEGKEILGTRKRGSP 93
 Db 5 LTKCOEVSHPNPAVHPSFRKCDENGNVLPDQYSGISGYCMCVFNGTEVNTSRGNH 64
 Oy 94 DCS 96
 Db 65 NCS 67

RESULT 14
 HLHMSG
 class II histocompatibility antigen-associated gamma chain II, 41k splice form -
 N:Alternate names: Ia-associated invariant chain II4; invariant chain proteoglyc
 N:Contains: class II histocompatibility antigen-associated gamma chain II, 31k sp
 C:Species: Mus musculus (house mouse)
 C:Date: 17-Mar-1987 #sequence_revision 12-Apr-1996 #text_change 22-Jun-1999
 C:Accession: B27866; A27866; A02244; A43530; S03099; S09284; A28956
 R:Koch, N.; Lauer, W.; Habicht, J.; Dobberstein, B.
 EMBO J. 6, 1677-1683, 1987
 A:Title: Primary structure of the gene for the murine Ia antigen-associated invar
 nce of thyroglobulin.
 A:Reference number: A91071; M0ID:87275861
 A:Accession: B27866
 A:Molecule type: DNA
 A:Residues: 1-279 <KOC>
 A:Experimental source: strain AKR
 A:Note: the authors translated the codon AAC for residue 70 as Asp
 A:Accession: A27866
 A:Molecule type: DNA
 A:Residues: 1-191,256-279 <KO2>
 A:Experimental source: strain AKR
 A:Note: the authors translated the codon AAC for residue 70 as Asp
 R:Singer, P.A.; Lauer, W.; Demblic, Z.; Mayer, W.E.; Lipp, J.; Koch, N.; Hammerling
 EMBO J. 3, 873-877, 1984
 A:Title: Structure of the murine Ia-associated invariant (II) chain as deduced fr
 A:Reference number: A02244; M0ID:84207946
 A:Accession: A02244
 A:Molecule type: mRNA
 A:Residues: 47-191,256-279 <STN>
 A:Cross-references: GB:X00496; NID:g53102; PIDN:CAA25191.1; PID:g53103
 R:Eades, A.M.; Liflin, M.; Rahmsdorf, H.J.
 J. Immunol. 144, 4399-4409, 1990
 A:Title: The IFN-gamma response of the murine invariant chain gene is mediated by
 A:Reference number: A43530; M0ID:90257363
 A:Accession: A43530
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-25 <EAD>
 A:Cross-references: GB:M35872; NID:g194129; PIDN:AAA37897.1; PID:g553934
 R:Stone, J.; Perry, R.; Todd, J.A.; McDevitt, H.O.

submitted to the EMBL Data Library, March 1988

A:Reference number: S03099

A:Accession: S03099

A:Molecule type: mRNA

A:Residues: 1-58 <STO>

A:Cross-references: EMBL:X07129; NID:952637; PIDN:CA30141.1; PID:952638

R:Zhu, L.; Jones, P.P.

Nucleic Acids Res. 17, 447-448, 1989

A:Title: Complete sequence of the murine invariant chain (II) gene.

A:Reference number: S09284; MUID:89098406

A:Accession: S09284

A:Molecule type: DNA

A:Residues: 1-9, 'M', 11-112, 'Q', 114-228, 'SG', 232-279 <ZHU>

R:Miller, J.; Hatch, J.A.; Simons, S.; Cullen, S.E.

Proc. Natl. Acad. Sci. U.S.A. 85, 1353-1353, 1988

A:Title: Identification of the glycosaminoglycan-attachment site of mouse invariant-chain

A:Reference number: A28956; MUID:88144436

A:Accession: A28956

A:Molecule type: mRNA

A:Residues: 261-270 <ML>

A:Note: Proof by site-directed mutagenesis that the sole chondroitin sulfate glycosaminoglycan

C/enetics:

A:introns: 26/2; 83/1; 109/3; 130/3; 161/3; 192/1; 256/1; 277/1

C:superfamily: class II histocompatibility antigen-associated gamma chain; thyroglobulin

C:keywords: alternative splicing; glycoprotein; transmembrane protein

F:1-279/Product: class II histocompatibility antigen-associated gamma chain II, 41k spl

F:1-191,256-279/Product: class II histocompatibility antigen-associated gamma chain II,

F:1-29/Domain: intracellular #status predicted <CYT>

F:30-55/Domain: transmembrane #status predicted <IMM>

F:56-279/Domain: extracellular #status predicted <EXT>

F:196-254/Domain: thyroglobulin type I repeat homology <THY1>

F:113,119,239,253/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:265/Binding site: carbohydrate (Ser) (covalent) #status experimental

Query Match 10.5%, Score 135, DB 1, Length 279;

Best Local Similarity 37.2%, Pred. No. 0.00034;

Matches 32; Conservative 11; Mismatches 35; Indels 8; Gaps 2;

QY 19 WLFELICATIT-----STENS---LTKCOQLQASANSGLIGTYVPQCKETGFEFEKQCGSGTGYCWCVDDEGKEIL 70

DB 170 WLFELKSKNSLEKKPTPEAPKVLTKQGEVSHIPAVYPGAFRKPCKDENGNTYLPLQCHGR 229

QY 71 TGTCWCVCVDDEGKEILGTRKSGSDCS 96

DB 230 HCYCWCVCVPNGTEVPTKSRGRNCS 255

RESULT 15

A45403

insulin-like growth factor-binding protein 4 - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 03-Dec-1999

C:Accession: A45403

R:Moser, D.R.; Lowe, J.R.; W.L.; Dake, B.L.; Booth, B.A.; Boes, M.; Clemmons, D.R.; Bar, F

Mol. Endocrinol. 6, 1805-1814, 1992

A:Title: Endothelial cells express insulin-like growth factor-binding proteins 2 to 6.

A:Reference number: A45403; MUID:93125553

A:Accession: A45403

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-258 <MOS>

A:Cross-references: GB:S52770; NID:9263303; PIDN:AA824873.1; PID:9263304

A:Experimental source: pulmonary artery endothelial cells

A:Note: sequence extracted from NCBI backbone (NCBI:122183, NCBI:122184)

C:superfamily: insulin-like growth factor binding protein 1; thyroglobulin type I repeat

F:174-249/Domain: thyroglobulin type I repeat homology <THY1>

Query Match 10.3%, Score 132.5; DB 2, Length 258;

Best Local Similarity 23.0%, Pred. No. 0.00081;

Matches 52; Conservative 33; Mismatches 70; Indels 71; Gaps 13;

QY 37 COQLQASANSGL-----IGTYVPQCKETGFEFEKQCGSGTGYCWCVDDEGKEIL 85
 DB 44 CEELVREPGGCGCATGALGKMPGCVYTPRCG-----SGLRCPYPRGV-----EKPL 90
 QY 86 GTRKSGSDCSRRKALITLCQMMQALITVNPVPCMGSPSCKADG-----SPDEVCCASNG 140
 DB 91 HTLVHGGGVCMELEIAIDQESIQ-----PSDKDGDHPNNSFS--PCSAHDR 136
 QY 141 ECYCVDR-----KGELEGTROQGRP-----TCERNHLSCEEARIKAHNSGLR 183
 DB 137 K--CLQKHLAKIKRDRSTSGGKMKVIGAPREAPVPQSCQSELHRALE-RLAASQSRTH 193
 QY 184 VEMF-----VPCLEDEGSTNPVQCPSPS-----TGTCWCVD-EGYKVPNG 221
 DB 194 EDLYTIPINCDNRGNFHPKQCHPALDGGRGKCMCVDRRTGVKLP 239

Search completed: March 8, 2002, 09:25:12
 Job time: 200 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2002, 09:21:32 ; Search time 12.57 Seconds

(without alignments)
413.545 Million cell updates/sec

Title: US-09-445-480A-2
Perfect score: 1287
Sequence: 1 MALSONQAKFSKGFVIMV.....VDEGVKVPBSPVRRKPRTC 231

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents,AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	122.5	9.5	186	1	US-08-698-551-6 Sequence 6, Appl1
2	122.5	9.5	186	2	US-08-602-228-6 Sequence 6, Appl1
3	122.5	9.5	186	2	US-08-649-341A-6 Sequence 6, Appl1
4	122.5	9.5	186	2	US-08-494-440B-6 Sequence 6, Appl1
5	122.5	9.5	186	2	US-08-533-901B-6 Sequence 6, Appl1
6	122.5	9.5	186	2	US-08-839-032A-6 Sequence 6, Appl1
7	122.5	9.5	186	2	US-08-839-031A-6 Sequence 6, Appl1
8	122.5	9.5	186	5	PCRT-US95-12724-6 Patent No. 5212074
9	122.5	9.5	237	6	US-08-698-551-8 Sequence 8, Appl1
10	122.5	9.5	272	2	US-08-602-228-8 Sequence 8, Appl1
11	122.5	9.5	272	2	US-08-649-341A-8 Sequence 8, Appl1
12	122.5	9.5	272	2	US-08-494-440B-8 Sequence 8, Appl1
13	122.5	9.5	272	2	US-08-533-901B-8 Sequence 8, Appl1
14	122.5	9.5	272	2	US-08-839-032A-8 Sequence 8, Appl1
15	122.5	9.5	272	2	US-08-839-031A-8 Sequence 8, Appl1
16	122.5	9.5	272	5	PCRT-US92-01196-4 Sequence 4, Appl1
17	122.5	9.5	272	5	PCRT-US95-12724-8 Sequence 8, Appl1
18	122.5	9.5	272	5	PCRT-US95-12724-8 Sequence 8, Appl1
19	118.5	8.8	271	5	PCRT-US92-01196-3 Patent No. 5212074
20	113.5	8.8	213	6	US-08-698-551-8 Sequence 8, Appl1
21	113.5	8.8	322	6	US-08-144-121-4 Sequence 4, Appl1
22	112	8.7	1196	2	US-08-735-893-4 Patent No. 5212074
23	112	8.7	1196	2	US-08-735-893-4 Patent No. 5212074
24	109.5	8.5	259	6	US-08-469-486-52 Sequence 52, Appl1
25	109.5	8.5	4544	1	US-08-469-558-52 Sequence 52, Appl1
26	109.5	8.5	4544	2	US-08-469-558-52 Sequence 52, Appl1
27	107	8.3	85	3	US-08-604-965E-9 Sequence 9, Appl1

28	107	8.3	264	1	US-08-482-271-3 Sequence 3, Appl1
29	107	8.3	264	1	US-08-482-271-4 Sequence 4, Appl1
30	107	8.3	264	2	US-08-854-811-45 Sequence 45, Appl1
31	107	8.3	264	3	US-09-080-120A-2 Sequence 2, Appl1
32	107	8.3	264	3	US-09-080-120A-4 Sequence 4, Appl1
33	107	8.3	264	5	PCRT-US95-08925-2 Sequence 2, Appl1
34	107	8.3	264	5	PCRT-US95-08925-2 Sequence 4, Appl1
35	107	8.3	261	1	US-08-468-847B-19 Sequence 19, Appl1
36	107	8.3	261	1	US-09-080-120A-7 Sequence 7, Appl1
37	107	8.3	261	5	PCRT-US95-08925-7 Patent No. 5212074
38	107	8.3	291	5	PCRT-US95-08925-7 Patent No. 5212074
39	106.5	8.3	292	6	US-08-820-170A-34 Sequence 34, Appl1
40	104	8.1	810	3	US-09-055-699-34 Sequence 34, Appl1
41	104	8.1	810	4	US-09-273-565-34 Sequence 34, Appl1
42	104	8.1	810	4	US-08-404-665-4 Sequence 4, Appl1
43	103.5	8.0	1358	1	US-08-404-671-4 Sequence 4, Appl1
44	103.5	8.0	1358	1	US-08-404-671-4 Sequence 4, Appl1
45	103.5	8.0	1358	1	US-08-404-781-4 Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-08-698-551-6
Sequence 6, Application US/08698551
Patent No. 5712381
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schiavella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNP RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/698,551
APPLICATION NUMBER: US/08/698,551
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-698-551-6

Query Match 9.5%; Score 122.5; DB 1; Length 186;
Best Local Similarity 35.1%; Pred. No. 3.4e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;
QY 151 ELECTNOGPTGERHLSDCEEARIKAHNSLRVEVFVPECELDGSSYNPVCQMPST---T 206
I: :||| |||| :||| :||| :|||

Db 96 EMOOSE0G--PCRHH-EASIOELKASPRMVPRAVYLPNCDRKGFKRKQCKPSRGRK 152
QY 207 GYCWCVDEGGVKKVPGSD 223
Db 153 GICWCVDKYGKMLPGME 169

RESULT 2

US-08-602-228-6
Sequence 6, Application US/08602228
Patent No. 5843675

GENERAL INFORMATION:

APPLICANT: Lin, Lin-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,228
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-602-228-6

Query Match

Best Local Similarity 9.5%; Score 122.5; DB 2; Length 186;
Best Local Similarity 35.1%; Pred. No. 3.4e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

QY 151 ELGTRGGRPTCEHRLSCFEARIKANSLSRVMEVPECTLEDGSTNPQCWPS----T 206
Db 96 EMOOSE0G--PCRHH-EASIOELKASPRMVPRAVYLPNCDRKGFKRKQCKPSRGRK 152
QY 207 GYCWCVDEGGVKKVPGSD 223
Db 153 GICWCVDKYGKMLPGME 169

RESULT 3

US-08-649-341A-6
Sequence 6, Application US/08649341A
Patent No. 5847099

GENERAL INFORMATION:

APPLICANT: Lin, Lin-Ling
APPLICANT: Chen, Jennifer H.
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/649,341A
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232-FWC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-649-341A-6

Query Match

Best Local Similarity 9.5%; Score 122.5; DB 2; Length 186;
Best Local Similarity 35.1%; Pred. No. 3.4e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

QY 151 ELGTRGGRPTCEHRLSCFEARIKANSLSRVMEVPECTLEDGSTNPQCWPS----T 206
Db 96 EMOOSE0G--PCRHH-EASIOELKASPRMVPRAVYLPNCDRKGFKRKQCKPSRGRK 152
QY 207 GYCWCVDEGGVKKVPGSD 223
Db 153 GICWCVDKYGKMLPGME 169

RESULT 4

US-08-494-440B-6
Sequence 6, Application US/0849440B
Patent No. 5849501

GENERAL INFORMATION:

APPLICANT: Lin, Lin-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/494,440B
FILING DATE:
CLASSIFICATION: 435

MOLECULE 11
US-08-533-901B-6

US-08-839-031A-6

MOLECULE 11
US-08-533-901B-6

US-08-839-031A-6

```

; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,031A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.,
; REGISTRATION NUMBER: 41,323
; REFERENCE/DOCKET NUMBER: G15232BDIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 186 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-839-031A-6

Query Match
Best local similarity 9.5%; Score 122.5; DB 2; Length 186;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

Qy 151 ELEGTRQGRPTCEHRLSECEARIRAHNSLREMFVEPECELEDGSGYNPQCWPS---T 206
| : | | | | | : | | : | : | | | | | | | |
Db 96 EMQSEEG--PCRHH-ESLQELKASPRMVPRAVYLPCNDKRGFKRKCKRSRGRK 152
Qy 207 GYCWCVDEGGYKVPQSD 223
| | | | | | | | | | | |
Db 153 GICWCVDKTKMKLPQME 169

RESULT 8
PCT-US95-12724-6
; Sequence 6, Application PC/TUS9512724
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/12724
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15232B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 186 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-12724-6

Query Match
Best local similarity 9.5%; Score 122.5; DB 5; Length 186;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

Qy 151 ELEGTRQGRPTCEHRLSECEARIRAHNSLREMFVEPECELEDGSGYNPQCWPS---T 206
| : | | | | | : | | : | : | | | | | | | |
Db 96 EMQSEEG--PCRHH-ESLQELKASPRMVPRAVYLPCNDKRGFKRKCKRSRGRK 152
Qy 207 GYCWCVDEGGYKVPQSD 223
| | | | | | | | | | | |
Db 153 GICWCVDKTKMKLPQME 169

RESULT 9
5212074-7
; Patent No. 5212074
; APPLICANT: KIEFER, MICHAEL C.; MASJAR, FRANK R.
; TITLE OF INVENTION: GENETIC MATERIAL ENCODING NEW
; INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN IGFBP-6
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/576,629
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 574,613
; FILING DATE: 28-AUG-1990
; SEQ ID NO: 7:
; LENGTH: 237
; 5212074-7

Query Match
Best local similarity 9.5%; Score 122.5; DB 6; Length 237;
Matches 52; Conservative 30; Mismatches 73; Indels 71; Gaps 13;

Qy 37 COQLQASANSGL-----IGTVYPOCKETGEFEDEKCKSGTYCWCVDDEKEL 85
| : | | | | | : | | : | : | | | | | | | |
Db 23 CEELVREPGCCATCALGSMFCGYTRCG-----SLKCIYPRGV-----EKPL 69
Qy 86 GTRKSGPDCSRRAALTLCCMMQALIVVPGWCGPSCKADG-----SFDEVQCCASNG 140
| : | | | | | : | | : | : | | | | | | | |
Db 70 HTLMGQGVCMELAEIAIQESIQ-----PSDKQGDHPNNSFS--PCSAHDR 115
Qy 141 EECVYDK-----KGKELGTRQGRP-----TCERHLSECEARIRAHNSLR 183
| : | | | | | : | | : | : | | | | | | | |
Db 116 R-CLQKHAKIRDRSTSGGKMKVNGAPREDAPVPQSCQSELHRALE-RLAASGRTH 172
Qy 184 VEMF---VPECELEDGSGYNPQCWPS---TYCCKVD-EGGYKVP 221
| : | | | | | : | | : | : | | | | | | | |
Db 173 EDLYTIPNCDRNGNFPKQCHPDLGGRGKWCVDKRTGKVLPG 218
```

```
RESULT 10
US-08-698-551-8
; Sequence 8, Application US/08698551
; Patent No. 5712381
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/698,551
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15232D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-698-551-8

Query Match          9.5%; Score 122.5; DB 1; Length 272;
Best Local Similarity 35.1%; Pred. No. 5.6e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

OY 151 ELESTRQGRPTCERHLSCECEARIRKHSNSLRVEMFVPCLEDGSYNPVQCWPS---T 206
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 182 EMROESSEG--PCRRHM-EASLDELKASPRMVPRAVYLPCNDRKGFYKRQCKPSRGRKR 238
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 207 GYCWCVDEGGYKVPYPSD 223
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 239 GICWCVDKYGKMLPGME 255

RESULT 11
US-08-602-228-8
; Sequence 8, Application US/08602228
; Patent No. 5843675
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
```

```
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,228
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15232C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-602-228-8

Query Match          9.5%; Score 122.5; DB 2; Length 272;
Best Local Similarity 35.1%; Pred. No. 5.6e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

OY 151 ELESTRQGRPTCERHLSCECEARIRKHSNSLRVEMFVPCLEDGSYNPVQCWPS---T 206
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 182 EMROESSEG--PCRRHM-EASLDELKASPRMVPRAVYLPCNDRKGFYKRQCKPSRGRKR 238
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 207 GYCWCVDEGGYKVPYPSD 223
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 239 GICWCVDKYGKMLPGME 255

RESULT 12
US-08-649-341A-8
; Sequence 8, Application US/08649341A
; Patent No. 5847099
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,341A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15232-FWC
```

```

? TELECOMMUNICATION INFORMATION
?
? TELEPHONE: (617) 498-8224
?
? TELEFAX: (617) 876-5651
?
? INFORMATION FOR SEQ ID NO: 8 :
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 272 amino acids
?
? type: amino acid
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: protein
?
? US-08-649-341A-8

```

Query Match	9.5%	Score 122.5;	DB 2;	Length 272;
Best Local Similarity	35.1%;	Pred. No. 5.6e-05;		
Matches 27, Conservative 12;	Mismatches 31;		Indels 7;	Gaps 3;

Dy 182 EMROSEEG - PCRHM- EASTJELKSPRNVRVAVLPLPNDRKGFGFKRRQCKPFSRKKR 238

Dc 207 GYCMCVDDGGGKYKVPDSD 223

Dd 239 GTCMCCVDATYGMKLPGME 255

RESULT 13
US-08-494-440B-8
; Sequence 8, Application US/08494440B
; Patent No. 5849501

APPLICANT: Lin, Lih-Ling
 APPLICANT: Chen, Jennifer H.
 APPLICANT: Schiavella, Andrea
 TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
 TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/94,440B
 FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: BROWN, Scott A,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G152324
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEO ID NO: 8:
SOURCES/CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-494-440B-8

Query Match 9.5%; Score 122.5; DB 2; Length 272;
Best Local Similarity 35.1%; Pred. No. 5.6e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps
0Y _q 151 ELECTROOGPTCCERHLSSECEARIKAHNSLAVEMFVPECLDGSYNPVQWPS---T 206

Db 182 ENRQEEGQ - PCKRH - EASLQELKSPRAYVAYLLPNCORKGFIYRKQCKPSRGRKR 236

QY 207 GYCMCVDGEGVYVPGSD 223
 1 11111 :1:111 :
 Db 239 GICMCVDYTKMKLPME 255

RESULT 14
US-08-533-901B-8
; Sequence 8, Application US/08533901E
; Patent No. 5853172

GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

Query Match	9.5%;	Score 122.5;	DB 2;	Length 272;
Best Local Similarity	35.1%;	Pred. No. 5.6e-05;		
Matches 27; Conservative	12;	Mismatches 31;	Indels 7;	Gaps 3

[illegible]

RESULT 15
 US-08-839-032A-8
 : Sequence 8, Application US/08839032A
 : Patent No. 5891675
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Iio, Iih-Ling
 : APPLICANT: Chen, Jennifer H.
 : APPLICANT: Schievella, Andrea

Mon Mar 11 08:00:35 2002

us-09-445-480a-2.rai

Page 7

```

APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,032A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15322BDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

```

	Query Match	9.5%	Score 122..5;	DB 2;	Length 272;
	Best Local Similarity	35.1%:	Pred. No. 5.6e-05;		
	Matches	27;	Conservative	12;	Mismatches 31; Indels 7; Gaps 3.
Oy	151	ELETRGCGPNCPEERLSCGEARIRKANSNSLRVEMFPECLEDGSGYNPOCPS----	T	206	
		: : : :			
Dd	182	EMRQESQG--PCRRRM-EASTLOELKASPRMYPRAVILPNDRCRGFTFKROCKPSKGRKR		238	
Oy	207	GTCMCVDDEGSVKRVGSD		223	
		: : :			
Dd	239	GICMCVDKRYGMKLPGME		255	

Search completed: March 8, 2002, 09:24:47
Job time: 195 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2002, 09:21:07 ; Search time 23.39 seconds

(without alignments)
731.549 Million cell updates/sec

Title: US-09-445-480a-2

Sequence: 1287
1 MALSQNAQAFKSGFYVMIV.....VDEGVKVPDGVFKKPRIC 231

Scoring table: BLOSUM62

Gapop 10.0, Capext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

A.Geneseq.1101.*
1: /SID58/gcgdata/geneseq/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1287	100.0	231	20	AA687988 Actinia equina L.
2	259.5	20.2	447	22	AA681111 Mouse TANGO 209 po
3	259.5	20.2	447	22	AA681442 Mouse TANGO 209 va
4	259.5	20.2	447	22	AA681443 Mouse TANGO 209 va
5	259.5	20.2	447	22	AA681444 Mouse TANGO 209 va
6	254.5	19.8	446	22	AA681307 Human TANGO 209 po
7	254.5	19.8	446	22	AA681339 Human TANGO 209 va
8	254.5	19.8	446	22	AA681401 Human TANGO 209 va
9	254.5	19.8	446	22	AA681411 Human TANGO 209 va
10	252.5	19.6	435	21	AA619394 Amino acid sequenc
11	247.5	19.2	238	21	AA656363 Human prostate can

12	221.5	17.2	176	22	AA625372 Human protein sequ
13	156	12.1	433	20	AA602602 HPTLG protein of h
14	156	12.1	436	19	AA674722 Human secreted pro
15	156	12.1	436	20	AA641753 Human PRO771 prote
16	156	12.1	436	21	AA641309 Human PRO771 (UNO4
17	156	12.1	436	21	AA624414 Human PRO771 prote
18	156	12.1	436	22	AA623906 Human EST encoded
19	156	12.1	436	22	AA624064 Human PRO771 poly
20	156	12.1	436	22	AA624064 Human gene 2 encod
21	156	12.1	447	22	AA603313 Human pancreatic c
22	153	11.9	422	21	AA654148 MHC Class II p41 s
23	150	11.7	296	20	AA631883 Human TANGO 209 po
24	148.5	11.5	45	22	AA648119 Bovine mammary tis
25	147.5	11.5	113	22	AA687613 Human immunostimul
26	141	10.3	45	22	AA648118 Human immunostimul
27	133	10.3	424	22	AA650906 Human PRO306 prote
28	133	10.0	233	21	AA672291 Rat insulin-like g
29	128.5	10.0	254	21	AA672290 Rat insulin-like g
30	128.5	10.0	233	21	AA608755 Human IGFBP-5
31	127.5	9.9	272	13	AA626995 Human IGFBP-5
32	125.5	9.8	113	20	AA625008 Insulin-like growt
33	125.5	9.5	186	17	AA635571 Tumor necrosis fac
34	122.5	9.5	186	18	AA635571 TNF-R1-UD ligand p
35	122.5	9.5	237	21	AA609620 Insulin-like growt
36	122.5	9.5	237	21	AA672933 Insulin-like growt
37	122.5	9.5	248	21	AA609756 IGFBP-5 amino acid
38	122.5	9.5	252	21	AA609621 Insulin like growt
39	122.5	9.5	258	13	AA622253 Sequence of insuli
40	122.5	9.5	258	13	AA621688 Insulin-like growt
41	122.5	9.5	258	21	AA672292 A human insulin-li
42	122.5	9.5	258	21	AA653968 Human PRO861 prote
43	122.5	9.5	258	22	AA650913 Human PRO861 prote
44	122.5	9.5	258	22	AA650913 Human PRO861 prote
45	122.5	9.5	258	22	AA650976 Human PRO861 prote

ALIGNMENTS

RESULT 1	AA687988	AA687988 standard; Protein; 231 AA.
ID	AA687988	
XX	AA687988	
AC	AA687988	
XX	09-APR-1999 (first entry)	
DT	XX	Actinia equina L. egulstatin protein.
DE	XX	Actinia equina L. egulstatin protein.
XX	XX	Actinia equina L. egulstatin protein.
KM	XX	type I repeated chrysothidin domain; cysteine protease inhibitor;
KM	XX	aspartic protease inhibitor; see anemone.
XX	XX	Actinia equina.
OS	XX	Actinia equina.
XX	XX	Actinia equina.
FT	XX	Key
FT	XX	Peptide
FT	XX	Location/Qualifiers
FT	XX	32..1
FT	XX	/label= signal
FT	XX	33..231
XX	XX	Protein
XX	XX	WO9858068-A2.
XX	XX	23-DEC-1998.
XX	XX	18-JUN-1998: 98WO-NL00352.
XX	XX	18-JUN-1997: 97EP-0201777.
XX	XX	(CPRO-) CPRO-DLO CENT PLANTENVERDEDELINGS REPROD.
XX	XX	Bosch HJ, Gruden K, Jongasma MA, Lenarcic B, Stiekema WJ;
XX	XX	Stukelj B, Turk V;
XX	XX	

XX 02-APR-2001 (first entry)
 DT
 XX
 DE Mouse TANGO 209 variant 1 polypeptide.
 XX
 KW TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
 KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
 KW antiasthmatic; neuroprotective; cytosolic; cardiant; hepatotropic;
 KW antiinflammatory; antidiabetic; antinfertility; antipyretic; vasotropic;
 KW antineumatic; nephrotropic; hemostatic; antilipemic; osteopathic;
 KW ophthalmological; antisticking; antitumor; vulnery; variant.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Msc-difference 65 /label= E65D
 FT /note= "wild-type Glu is replaced by Asp"
 XX
 PN W0200069885-A2.
 XX
 PD 23-NOV-2000.
 XX
 PF 15-MAY-2000; 2000MO-US13361.
 XX
 PR 14-MAY-1999; 99US-0312359.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Pan Y, Leiby KR;
 XX
 DR WPI: 2001-024999/03.
 DR N-PSDB: AAC84410.
 XX
 PT Novel nucleic acids encoding secreted or transmembrane proteins, useful
 PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
 PT the lung, liver, kidney or pancreas -
 XX
 PS Claim 8; Page -: 209pp; English.
 XX
 CC The invention provides human and mouse nucleic acids designated TANGO
 CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
 CC proteins. The polypeptides, nucleic acids and their modulators may be
 CC useful for treating or modulating cholesterol uptake, blood coagulation,
 CC to modulate cell proliferation, morphogenesis and fate specification,
 CC tissue repair and renewal, to treat cancer and promote wound healing,
 CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
 CC syndrome, protein S deficiency, modulate allergic or inflammatory
 CC response, acid secretion, tropic effects on gastrointestinal mucosa, and
 CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
 CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
 CC herniations, meningitis, ischemic brain or heart disease, infarction,
 CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
 CC disease, pulmonary heart disease, rheumatic fever, congenital heart
 CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
 CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
 CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
 CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, Mcune
 CC Albright syndrome, infertility, uterine disorders, viral disease. The
 CC present sequence represents a mouse TANGO 209 variant polypeptide.
 CC Note: the present variant sequence has been constructed using the
 CC information provided in the specification.
 CC
 XX Sequence 447 AA:
 SO

Query Match 20.2%; Score 259.5; DB 22; Length 447;
 Best Local Similarity 28.9%; Pred. No. 1.3e-14;
 Matches 36; Conservative 30; Mismatches 86; Indels 27; Gaps 5;

41 QASANSGLIGTYVPOCKTEGFEERKOCWGSTGYCWCVDDEKELITKRI-RGSPDCSRK 99

Db 98 qeakarkefqvflpecdnddgtysqyqchsygycwctvtngrpsstgavahkxprp---- 154
 Oy 100 AALTLCOMAOITVWVPCWCG-----PPSCRADSFDEVCCASNGCYCVDKKKGE 151
 Db 155 -----gslnexvptregagkaddaapaletpqgdeedias-typlwtegyksrq 205
 Oy 152 LEEGRDQGRPTCERHLSDEEFARIKAHNSLRVEMFVPECLDEDSYVPOGWPSTGYCWC 211
 Db 206 ntknksaascdeghsalleakqpkndh-----vvlpeahgyllykpvqchpsstgycwc 260
 Oy 212 V-DEGCVKVPGSVPRKRTYC 231
 Db 261 vlvdtgrplpqtsttyeqpkc 281

RESULT 4
 AAB48143
 ID AAB48143 standard; Protein; 447 AA.
 AC AAB48143;
 XX
 DT 02-APR-2001 (first entry)
 XX
 DE Mouse TANGO 209 variant 2 polypeptide.
 XX
 KW TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
 KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
 KW antiasthmatic; neuroprotective; cytosolic; cardiant; hepatotropic;
 KW antiinflammatory; antidiabetic; antinfertility; antipyretic; vasotropic;
 KW antineumatic; nephrotropic; hemostatic; antilipemic; osteopathic;
 KW ophthalmological; antisticking; antitumor; vulnery; variant.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Msc-difference 77 /label= E77D
 FT /note= "wild-type Glu is replaced by Asp"
 XX
 PN W0200069885-A2.
 XX
 PD 23-NOV-2000.
 XX
 PF 15-MAY-2000; 2000MO-US13361.
 XX
 PR 14-MAY-1999; 99US-0312359.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Pan Y, Leiby KR;
 XX
 DR WPI: 2001-024999/03.
 DR N-PSDB: AAC84411.
 XX
 PT Novel nucleic acids encoding secreted or transmembrane proteins, useful
 PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
 PT the lung, liver, kidney or pancreas -
 XX
 PS Claim 8; Page -: 209pp; English.
 XX
 CC The invention provides human and mouse nucleic acids designated TANGO
 CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
 CC proteins. The polypeptides, nucleic acids and their modulators may be
 CC useful for treating or modulating cholesterol uptake, blood coagulation,
 CC to modulate cell proliferation, morphogenesis and fate specification,
 CC tissue repair and renewal, to treat cancer and promote wound healing,
 CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
 CC syndrome, protein S deficiency, modulate allergic or inflammatory
 CC response, acid secretion, tropic effects on gastrointestinal mucosa, and
 CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
 CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain

KW ophthalmological; antislackling; antilicer; vulnerary.

XX Homo sapiens.

XX WO200069885-A2.

XX 23-NOV-2000.

XX 15-MAY-2000; 2000WO-US13361.

XX 14-MAY-1999; 99US-0312359.

XX (MILL-) MILLENNIUM PHARM INC.

PA Pan Y, Leiby KR;

PI WPI: 2001-024999/03.

DR N-PSDB; AAC84381, AAC84382.

XX Novel nucleic acids encoding secreted or transmembrane proteins, useful for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of the lung, liver, kidney or pancreas -

XX Claim 8; Fig 14A-E; 209pp; English.

XX The invention provides human and mouse nucleic acids designated TANGO 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane proteins. The polypeptides, nucleic acids and their modulators may be useful for treating or modulating cholesterol uptake, blood coagulation, to modulate cell proliferation, morphogenesis and fate specification, tissue repair and renewal, to treat cancer and promote wound healing, modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan syndrome, protein S deficiency, modulate allergic or inflammatory response, acid secretion, tropic effects on gastrointestinal mucosa, and promote ulcer healing, treat bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis, leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain herniations, meningitis, pancreatitis, diabetes, angina, hypotensive heart disease, intracranial hemorrhage, pulmonary heart disease, congenital heart disease, myocardial disease, atherosclerosis, hypertension, jaundice, hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome, sickle cell disease, renal failure, ischemic bowel disease, Crohn's disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome, CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune CC Albright syndrome, infertility, uterine disorders, viral disease. The CC present sequence represents the human TANGO 209 polypeptide.

XX Sequence 446 AA:

XX Query Match 19.8%; Score 254.5; DB 22; Length 446;

XX Best Local Similarity 28.2%; Pred. No. 3,6e-14;

XX Matches 59; Conservative 29; Mismatches 78; Indels 43; Gaps 7;

XX 41 QASANSGLITGVPOCKEETGEFEKOCWSTGYCWCVDDEGEILGRTI-RGSPDCRRK 99

XX 98 gqgarketqyffipenddgtysqyqchsytygcwcvupngpislgtavakprc----- 153

XX 100 AALTLGQMMAALIVNPGMGP--PSCKADGSPDEV-----OCCANSGCYC 144

XX 154 -----pysvneklpqreigtgtdaapaletqpgqdeaiasryptlw 137

XX 145 VVKKKKELEGRROGRTPTCER-HLSECEARIKAHNSIRKEMVPECLDEGSLNPOCW 203

XX 198 teqyvksgnktnksvsscdqeqsaeekqpkndh-----vvlpecatggylykpvqch 252

XX 204 PSTGYCMCV-DEGGKVVPGSDVFRKRPYC 231

XX 253 pstgywcvlvtgtrpigtsttryepkpc 281

RESULT 7

AAB48139

XX ID AAB48139 standard; Protein: 446 AA.

XX AAB48139;

XX 02-APR-2001 (first entry)

XX Human TANGO 209 variant 1 polypeptide.

XX TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;

XX transmembrane protein; antineutrophilic; cerebroprotective; arteriosclerosis;

XX antiasthmatic; neutrophilic; cytosolic; cardiant; hepatotropic;

XX antiinflammatory; antidiabetic; antinfertility; antipyloric; vasotrophic;

XX antirheumatic; nephrotropic; hemostatic; antileptic; osteopachic;

XX ophthalmological; antislackling; antilicer; vulnerary; variant.

XX Homo sapiens.

XX Key

XX Misc-difference 65

XX /label= E65D

XX /note="wild-type Glu is replaced by Asp"

XX WO200069885-A2.

XX 23-NOV-2000.

XX 15-MAY-2000; 2000WO-US13361.

XX 14-MAY-1999; 99US-0312359.

XX (MILL-) MILLENNIUM PHARM INC.

XX Pan Y, Leiby KR;

XX WPI: 2001-024999/03.

XX N-PSDB; AAC84407.

XX Novel nucleic acids encoding secreted or transmembrane proteins, useful for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of the lung, liver, kidney or pancreas -

XX Claim 8; Page -; 209pp; English.

XX The invention provides human and mouse nucleic acids designated TANGO 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane proteins. The polypeptides, nucleic acids and their modulators may be useful for treating or modulating cholesterol uptake, blood coagulation, to modulate cell proliferation, morphogenesis and fate specification, tissue repair and renewal, to treat cancer and promote wound healing, modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan syndrome, protein S deficiency, modulate allergic or inflammatory response, acid secretion, tropic effects on gastrointestinal mucosa, and promote ulcer healing, treat bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis, leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain herniations, meningitis, pancreatitis, diabetes, angina, hypotensive heart disease, intracranial hemorrhage, pulmonary heart disease, congenital heart disease, myocardial disease, atherosclerosis, hypertension, jaundice, hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome, sickle cell disease, renal failure, ischemic bowel disease, Crohn's disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome, CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune CC Albright syndrome, infertility, uterine disorders, viral disease. The CC present sequence represents a human TANGO 209 variant polypeptide. Note: the present variant sequence has been constructed using the information provided in the specification.

XX Sequence 446 AA:

XX Query Match 19.8%; Score 254.5; DB 22; Length 446;

Best Local Similarity 28.2%; Pred. No. 3.6e-14;
Matches 59; Conservative 29; Mismatches 78; Indels 43; Gaps 7;

```

QY 41 QASANSGLIGYVPCKEETGEFEKQWSTGYCWCVDDEGKEILGTFI-RGSPDCSRK 99
DB 98 qegetketqyflpeccnddgtysqvgchytgywcvcvprpplsgtavahktpc----- 153
QY 100 AALILCOMQAIIVNPGWCGP--PSCKADGSPDEV-----OCCASNGECYC 144
DB 154 -----psvneklpqregtgytdaapaletcpqgdeedaaryptlw 197
QY 145 VDKKGELEGTROOGRPTCER-HLSECEARIKAHNSLRVEMFPECELEGSYNPQWCW 203
DB 198 tegvksrqnkukhsvescdqeqbsaleekqpkndn-----vvipecahggllykpvqch 252
QY 204 PSTGYCMCV-DEGSVKVPGSDVFRKRPCTC 231
DB 253 pstgywcvcvldtgrpdpstctryeqpkc 281

```

RESULT 8
AAB48140
ID AAB48140 standard; Protein: 446 AA.

AC AAB48140;
DT 02-APR-2001 (first entry)

DE Human TANGO 209 variant 2 polypeptide.

XX TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
KW transmembrane protein; anitaneimic; cerebroprotective; arteriosclerosis;
KW antiasthmatic; neuroprotective; cytoskeletal; cardiant; hepatotropic;
KW antiinflammatory; antidiabetic; antifertility; antipyretic; vasotropic;
KW antirheumatic; nephrotropic; hemostatic; antilipemic; osteopathic;
KW ophthalmological; antislackling; antilucer; vulnerary; variant.

OS Homo sapiens.

Key Location/Qualifiers
FH Misc-difference 77 /label= E77D
FT /note= "wild-type Glu is replaced by Asp"

XX WO200069885-A2.
XX 23-NOV-2000.

XX 15-MAY-2000; 2000MO-US13361.
XX 14-MAY-1999; 99US-0312359.

XX (MILL-) MILLENNIUM PHARM INC.
XX Pan Y, Leiby KR;

XX WPI: 2001-02A999/03.
XX N-PSDB; AAC844408.

XX Novel nucleic acids encoding secreted or transmembrane proteins, useful
XX for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
XX the lung, liver, kidney or pancreas -

XX Claim 8; Page - 209pp; English.

XX The invention provides human and mouse nucleic acids designated TANGO
XX 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
XX proteins. The polypeptides, nucleic acids and their modulators may be
XX useful for treating or modulating cholesterol uptake, blood coagulation,
XX to modulate cell proliferation, morphogenesis and fate specification,
XX tissue repair and renewal, to treat cancer and promote wound healing,
XX modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
XX syndrome, protein S deficiency, modulate allergic or inflammatory

CC response, acid secretion, tropic effects on gastrointestinal mucosa, and
CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
CC herniations, meningitis, ischemic brain or heart disease, infarction,
CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
CC disease, pulmonary heart disease, rheumatic fever, congenital heart
CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
CC Albright syndrome, infertility, uterine disorders, viral disease. The
CC present sequence represents a human TANGO 209 variant polypeptide. The
CC Note: the present variant sequence has been constructed using the
XX information provided in the specification.

Sequence 446 AA;

Query Match 19.8%; Score 254.5; DB 22; Length 446;
Best Local Similarity 28.2%; Pred. No. 3.6e-14;
Matches 59; Conservative 29; Mismatches 78; Indels 43; Gaps 7;

```

QY 41 QASANSGLIGYVPCKEETGEFEKQWSTGYCWCVDDEGKEILGTFI-RGSPDCSRK 99
DB 98 qegetketqyflpeccnddgtysqvgchytgywcvcvprpplsgtavahktpc----- 153
QY 100 AALILCOMQAIIVNPGWCGP--PSCKADGSPDEV-----OCCASNGECYC 144
DB 154 -----psvneklpqregtgytdaapaletcpqgdeedaaryptlw 197
QY 145 VDKKGELEGTROOGRPTCER-HLSECEARIKAHNSLRVEMFPECELEGSYNPQWCW 203
DB 198 tegvksrqnkukhsvescdqeqbsaleekqpkndn-----vvipecahggllykpvqch 252
QY 204 PSTGYCMCV-DEGSVKVPGSDVFRKRPCTC 231
DB 253 pstgywcvcvldtgrpdpstctryeqpkc 281

```

RESULT 9
AAB48141
ID AAB48141 standard; Protein: 446 AA.

AC AAB48141;

DT 02-APR-2001 (first entry)

DE Human TANGO 209 variant 3 polypeptide.

XX TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
KW transmembrane protein; anitaneimic; cerebroprotective; arteriosclerosis;
KW antiasthmatic; neuroprotective; cytoskeletal; cardiant; hepatotropic;
KW antiinflammatory; antidiabetic; antifertility; antipyretic; vasotropic;
KW antirheumatic; nephrotropic; hemostatic; antilipemic; osteopathic;
KW ophthalmological; antislackling; antilucer; vulnerary; variant.

OS Homo sapiens.

Key Location/Qualifiers
FH Misc-difference 93 /label= E93D
FT /note= "wild-type Glu is replaced by Asp"

XX WO200069885-A2.
XX 23-NOV-2000.

XX 15-MAY-2000; 2000MO-US13361.
XX 14-MAY-1999; 99US-0312359.

PA (MILL-) MILLENNIUM PHARM INC.
 XX Pan Y, Leiby KR;
 XX WPI: 2001-024999/03.
 DR N-PSDB: AAC84409.
 XX Novel nucleic acids encoding secreted or transmembrane proteins, useful
 PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
 PT the lung, liver, kidney or pancreas.
 XX
 PS Claim 8; Page -: 209pp; English.
 XX
 CC The invention provides human and mouse nucleic acids designated TANCO
 CC 204, TANCO 206, TANCO 209 and A236 encoding secreted or transmembrane
 CC proteins. The polypeptides, nucleic acids and their modulators may be
 CC useful for treating or modulating cholesterol uptake, blood coagulation,
 CC to modulate cell proliferation, morphogenesis and fate specification,
 CC tissue repair and renewal, to treat cancer and promote wound healing,
 CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
 CC syndrome, protein S deficiency, modulate allergic or inflammatory
 CC response, acid secretion, tropic effects on gastrointestinal mucosa, and
 CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
 CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
 CC infections, meningitis, ischemic brain or heart disease, infarction,
 CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
 CC disease, pulmonary heart disease, rheumatic fever, congenital heart
 CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
 CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
 CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
 CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McBurney
 CC Albright syndrome, infertility, uterine disorders, viral disease. The
 CC present sequence represents a human TANCO 209 variant polypeptide.
 CC Note: the present variant sequence has been constructed using the
 CC information provided in the specification.
 XX
 SQ Sequence 446 AA:
 Query Match 19.8%; Score 254.5; DB 22; Length 446;
 Best Local Similarity 28.2%; Pred. No. 3, 6e-14;
 Matches 59; Conservative 29; Mismatches 78; Indels 43; Gaps 7;
 QY 41 OASANSGLIGTYVPOCKETGEFEKOCMSGYCWCVDDEGKEITGTI-NGSPCKSRK 99
 DB 98 qegarketqyvfpecodgdtysqvgchstygyewcpvprgpiqstevahktprc----- 153
 QY 100 AALTLCOMQAIIVNPGWGP--PSCKADGSPDEV-----GCCASNGECYC 144
 DB 154 -----pgsvneklpqregtgktdaaapaletqpgqdeedlastrytliw 197
 QY 145 VDKKKELEGTROOGRPTCR-HLSECEARIKAHNSLRVEMFPECELDGSSYPVQCM 203
 DB 198 tegvysygnktknksvsodqegnsaleekqkpnkn-----vvlpecahgilykprvch 252
 QY 204 PSTGTGCV-DEGCVKVGSDVRKRPCTC 231
 DB 253 pstgywcwlvdtgrpigrpstrtyeqpkc 281
 RESULT 10
 AAB19394
 ID AAB19394 standard; Protein: 435 AA.
 AC AAB19394;
 XX
 XX 06-MAR-2001 (first entry)
 XX
 XX Amino acid sequence of a human secreted protein.
 DE
 XX Secreted protein; platelet disorder; stem cell disorder; osteoporosis;
 KW

KW osteoarthritis; burn; incision; ulcer; periodontal disease; neuropathy;
 KW nervous system disease; bone growth; cosmetic plastic surgery;
 KW gut protection; gut regeneration; fibrosis; cancer;
 KW bone marrow transplantation.
 XX
 XX Homo sapiens.
 OS
 XX WO200061755-A2.
 PN
 XX 19-OCT-2000.
 PD
 XX 10-APR-2000; 2000MO-US09555.
 PF
 XX
 PR 09-APR-1999; 99US-0128574.
 PR 20-AUG-1999; 99US-0150054.
 XX
 PA (CHIR) CHIRON CORP.
 XX Garcia PD;
 DR WPI: 2000-665133/64.
 DR N-PSDB: AAC61880.
 XX
 PT Novel secreted human proteins useful for stimulating blood cell
 PT generation in patients receiving cancer chemotherapy, treating bone
 PT marrow transplantation patients and for healing fractured bones.
 XX
 PS Claim 14; Page 55; 74pp; English.
 XX
 XX AAB19393-B19407 represent secreted human proteins. The secreted proteins
 XX are useful in assays to determine their biological activities. The
 XX proteins can also be used as biomarkers to identify tissues or cell
 XX types which express the proteins. The polynucleotide molecules can be
 XX used as biomarkers for tissues or chromosomes and to elicit immune
 XX responses. The proteins and antibodies are useful in diagnosis and
 XX treatment of diseases associated with altered expression of these
 XX proteins. The proteins are also useful for prevention or treatment of
 XX platelet disorders, stem cell disorders, osteoporosis or osteoarthritis,
 XX burns, incisions, ulcers, periodontal diseases, central and peripheral
 XX nervous system diseases and neuropathies, for healing fractured bones
 XX and to induce cartilage and/or bone growth in cosmetic plastic surgery.
 CC The proteins are also useful for gut protection or regeneration, for the
 CC treatment of lung or liver fibrosis, for stimulating blood cell
 CC generation in patients receiving cancer chemotherapy and for treatment
 CC of bone marrow transplantation patients.
 XX
 SQ Sequence 435 AA:
 Query Match 19.68%; Score 252.5; DB 21; Length 435;
 Best Local Similarity 31.7%; Pred. No. 5, 3e-14;
 Matches 70; Conservative 26; Mismatches 88; Indels 37; Gaps 8;
 QY 31 EASLTKCO-----QOASANSGLIGTYVPOCKETGEFEKOCMSGYCWCVDDEGKEITG 86
 DB 89 dagqskcrleraqlaqgakkpgevtfpecgdsstlyqchtygyewcpvprgpiqstevahktprc 148
 QY 87 TKIRG-SPDCSRRAAALTLCOMQAIIVNPGWGP--PSCKADGSPDEV-----FDVQ 134
 DB 149 ssvgnktktpvcsvldkplsg-----gnsgrkddqskpctmetqpyfdgde 195
 QY 135 CCASNGECYCVDKKGKLEGTROOGRP--TCERHLSCECEARIKAHNSLRVEMFPECELDGSSYPVQCM 191
 DB 196 lkaplwlkhlvlkdsklmnltnlnsekvyscdq-----ergsalleeqnpr-regtlvpec 251
 QY 192 LEDGSYNVQCMPTSTGYCWCY-DEGCVKVGSDVRKRPCTC 231
 DB 252 apgilykprvqchgstgyewcvlvtgrpigrpstrtyeqpkc 292
 RESULT 11
 AAB56363
 ID AAB56363 standard; Protein: 238 AA.
 KW

XX AAB56363;
XX 13-MAR-2001 (first entry)
XX Human prostate cancer antigen protein sequence SEQ ID NO:941.
XX
XX
XX
XX Human prostate cancer; prostate cancer antigen; detection; diagnosis;
XX neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;
XX vulnery; gastrointestinal; nephrotoxic; antiinfective; gynaecological;
XX antibacterial; gene therapy; neural; immune; reproductive; renal;
XX gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
XX wound; infectious disease.
XX
XX Homo sapiens.
XX WO20055174-A1.
XX 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US05988.
XX 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-587513/55.
XX N-PSDB: AAF15566.
XX
XX Prostate cancer associated gene sequences, referred to as prostate
XX cancer antigens, useful for treatment, prevention, and diagnosis of
XX disorders such as prostate cancer -
XX
XX Claim 11; Page 1381; 2338pp; English.
XX
XX AAF15566 to AAF16505 encode the human prostate cancer associated
XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
XX The prostate cancer antigens can have neuroprotective, cytosolic,
XX cardioactive, immunomodulatory, muscular, vulnery, gastrointestinal,
XX nephrotoxic, antiinfective, gynaecological and antibacterial activities,
XX and can be used in gene therapy. The prostate cancer antigen
XX polynucleotides may be used for detection of prostate cancer, chromosome
XX identification, as chromosome markers, and for numerous other diagnostic
XX or research purposes. The prostate cancer antigens may be used to treat
XX disorders such as neural, immune, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
XX AAB57303 represent sequences used in the exemplification of the present
XX invention.
XX
XX Sequence 238 AA:
SQ
Query Match 19.2%; Score 247.5; DB 21; Length 238;
Best Local Similarity 30.0%; Pred. No. 7e-14;
Matches 63; Conservative 24; Mismatches 62; Indels 61; Gaps 7;
DB 27 ITTSEALTKCQOLQ--ASANSGLIGT--VYPOCKEGERPEKOCWSTGYCWCVEDGK 82
DB 56 lpdstsltpceqqrhagayayparfhipqcdqgnlilpjqchgstqfvcwvdpdgh 115
DB 83 EILGTKIRGSPDCSRRAALTLICOMOAIIIVNPGWGPCPSCKADGSFDEVCASNGEC 142
DB 116 evpqlt-----lppgstlphhc----- 131
DB 143 YCVDKKKELEGTROGRPT-CERHLSCEBARIKAHNSLRVEMVPECLDGSYVPVQ 201
DB 132 -----gspsept--qrpptlcer-----wrenllehygsprddqyvpqcdldghlfpj 179
DB 202 CWPSTGYCWCVDGSGVAVPGSDVAFKRPKC 231

DB 180 chgksdftcwcvdkgrevvgtgxparphpc 209
RESULT 12
ID AAM25372
ID AAM25372 standard; Protein; 176 AA.
XX
XX AAM25372;
XX 16-OCT-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:887.
XX
XX Human: cancer; ulcer; HIV infection; human immunodeficiency virus;
XX antinflammatory; antirheumatic; antiarthritic; immunosuppressive;
XX antibacterial; endocrine; cardiant; central nervous system; vitruclide;
XX anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;
XX antiagregant; haemostatic; vulnery; antidiabetic; cytosolic;
XX dermatological; anti-allergic; antistimatic; antidiabetic; cytosolic;
XX neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
XX immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
XX anaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
XX cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
XX genetic disease; haematopoietic disorder; platelet disorder; asthma;
XX thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
XX allergic rhinitis; diabetes; multiple sclerosis; depression;
XX Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
XX neurological disorder.
XX
XX Homo sapiens.
XX
XX WO200153455-A2.
XX 26-JUL-2001.
XX
XX 22-DEC-2000; 2000WO-US35017.
XX
XX 23-DEC-1999; 99US-0471275.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Dimaac RT;
XX
XX WPI: 2001-457603/49.
XX N-PSDB: AAH99313.
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
XX Claim 20; Page 196; 1217pp; English.
XX
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
XX AAM25963. The proteins can have activities based on the tissues and
XX cells they are expressed in, such as: antinflammatory; antirheumatic;
XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
XX central nervous system; vitruclide; anti-HIV; fungicide; antimutagen;
XX cardiovascular; antianemic; antiagregant; haemostatic; vulnery;
XX antidiabetic; osteopathic; dermatological; anti-allergic; antistimatic;
XX antiparkinsonian; gene therapy; antisense therapy; vaccine and vaccine
XX encoding them can be used in gene therapy, antisense therapy and vaccine
XX production. The proteins and polynucleotides are useful for screening for
XX agonists or antagonists of a protein and for the treatment and diagnosis
XX of disorders associated with the activity of a protein e.g. Inflammation,
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XX infections, autoimmunity, genetic diseases, haematopoietic disorders,
XX anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
XX osteoporosis, severe combined immunodeficiency, eczema, allergic
XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,

AA The present invention describes secreted and transmembrane polypeptides
CC and their polynucleotides. The nucleotide sequences are useful as
CC

sources of probes, primers, for chromosome mapping, and for genetic
artefact sequences. They can be used to create transgenic
animals. The proteins can be used to treat a variety of diseases and
disorders, depending on their function. Diseases that may be treated
include blood coagulation disorders, cancers and cellular adhesion
disorders. They may also be used to raise antibodies. AA23891 to
AA34338, and AA14685 to AA14774 represent polynucleotide and
polypeptide sequence given in the exemplification of the present
invention.

50 Sequence 436 AA;

Query Match	12.18;	Score 156;	DB 20;	Length 436;
Best Local Similarity	40.98;	Pred No. 1	3e-05;	

Best Local Similarity 40.9%; Pred. No. 1.3e-05;
Matches 27; Conservative 12; Mismatches 25; Indels 2; Gaps 1;

Matches 27; Conservative 12; Mismatches 25; Indels 2; Gaps 1;

QY 31 EASLTKCQQLQASANSGLIGTYVPQCKETGFEFEKQCWGSTGYCWCVDEDEGKEILGTKIR 90

Db 318 qtelstnqkrqgvkk--llgqylplcdedgyykptqchsgsvqgcwcdrygnevmgsrrln 375

QY 91 GSPDCS 96

Db 376 gvađca 381

Search completed: March 8, 2002, 09:24:28
Job time: 201 sec

1
2
3
4